Revisiting Targeting in Social Assistance

HUMAN DEVELOPMENT PERSPECTIVES

Revisiting Targeting in Social Assistance:

A New Look at Old Dilemmas

Margaret Grosh, Phillippe Leite, Matthew Wai-Poi, and Emil Tesliuc, Editors

WORLD BANK GROUP

Human Development Perspectives

The books in this series address main and emerging development issues of a global/regional nature through original research and findings in the areas of education, gender, health, nutrition, population, and social protection and jobs. The series is aimed at policy makers and area experts and is overseen by the Human Development Practice Group Chief Economist.

Previous titles in this series

Feng Zhao, Clemens Benedikt, and David Wilson, *Tackling the World's Fastest-Growing HIV Epidemic: More Efficient HIV Responses in Eastern Europe and Central Asia* (2020).

Meera Shekar and Barry Popkin, *Obesity: Health and Economic Consequences of an Impending Global Challenge* (2020).

Truman Packard, Ugo Gentilini, Margaret Grosh, Philip O'Keefe, Robert Palacios, David Robalino, and Indhira Santos, *Protecting All: Risk Sharing for a Diverse and Diversifying World of Work* (2019).

Damien de Walque, *Risking Your Health: Causes, Consequences, and Interventions to Prevent Risky Behaviors* (2014).

Rita Almeida, Jere Behrman, and David Robalino, *The Right Skills for the Job? Rethinking Training Policies for Workers* (2012).

Barbara Bruns, Deon Filmer, and Harry Anthony Patrinos, *Making Schools Work: New Evidence on Accountability Reforms* (2011).

Harold Alderman, *No Small Matter: The Impact of Poverty, Shocks, and Human Capital Investments in Early Childhood Development* (2011).

All books in the Human Development Perspectives series are available at https://openknowledge.worldbank.org/handle/10986/2161.

Contents

Chapter 2: **Unpacking the Empirics of Targeting in
Low- and Middle-Income Countries** **81**

*Margaret Grosh, Claudia P. Rodriguez Alas, Usama Zafar,
Matthew Wai-Poi, Emil Tesliuc, and Phillippe Leite*

Figures

Maps

Photos

Tables

Foreword

Universal social protection is a goal shared by governments around the world, codified in the Sustainable Development Goals, and supported by many international partners in development, including the World Bank. Social protection policies and programs help individuals and societies manage risk and mitigate poverty through instruments that improve resilience, equity, and opportunity.

One of the most important decisions for policy makers in all countries is the differentiation of eligibility and benefits among people, commonly known in social protection as prioritization or targeting. Targeting is a common element of social protection policies and programs. Especially in countries with the highest coverage, social protection is built from a series of programs of different sorts. Some programs pay out depending on the state of the worker (for example, unemployment, retirement, illness, or disability) or individual (for example, age), and some pay out depending on the state of the family (for example, contingent on poverty or incurring a loss due to a natural disaster). Targeting is a useful tool as societies seek to fulfill the moral imperative of helping the neediest first or most.

The task of prioritizing among individuals or groups—or targeting—is fraught with conceptual and practical difficulties, with errors and costs, and with many criteria and metrics by which success or lack thereof can be gauged. Thus, debates recur around decisions about how broadly or narrowly to target, the choice of targeting method, and the many ways that a method can be customized to a specific setting. This book focuses on these themes, examining both principles and practice.

The book focuses on programs that intend to differentiate eligibility or benefit levels along the spectrum of money-metric welfare and the shocks that affect it. It looks at methods that measure or infer household welfare and methods that use more aggregated geographic or demographic categories that are expected to be correlated with money-metric poverty. The idea of targeting applies more broadly than only with respect to money-metric deprivation. It also includes consideration of who is more in need of policy support in various ways (for example, to improve the lives of those coping with a disability, or those lagging in employment or education, or those coping with other challenges, including climate-related ones). Although there are several parallels or analogues between money-metric and other types of targeting—especially in thinking about tolerance for inclusion or exclusion errors, the need to build good administrative systems, and concern about their cost, the need to consider the human rights and behaviors of those involved—the tools of discernment among poverty, disability, employability, and other factors are different. Accordingly, to keep the topic manageable, the book focuses on programs related to money-metric welfare.

In many ways, the key takeaways about targeting in social assistance are perennial. The core set of methods remains the same, good data and systems remain critical, country context remains paramount, and implementation matters. A rich new practice, however, shows how innovation or diligently building capacity can overcome challenges, how practical and important customizations to the setting can be done, how different desirable features in the targeting process can be in tension with each other, and how new data and technology may be changing the field.

There is no single best way of targeting for all circumstances. Moreover, different approaches may have different appeals in the same context. The book should not be read as a manual but rather as a guide to thinking through key questions. What are the policy objectives for a particular program? What is the capacity centrally and locally, and how can it be augmented? What data are available or can be easily obtained? What counts as success? The answers to these questions will influence the choices about targeting. The book advocates a robust process for arriving at a context-specific solution rather than adopting a generic good practice.

The COVID-19 (coronavirus) pandemic and the social and economic chaos it has wrought globally continue to hang over the world. Government responses to the crisis have, to a great degree, determined how well households have borne up or sunk down. Whether there has been enough assistance to the right people has been key; decisions about when to taper assistance, and for whom, will shape the path of recovery for families and nations. At the same time, COVID-19 has highlighted the importance of social protection systems that can respond quickly to shocks, assisting the newly vulnerable as well as those already in need. COVID-19 will fade, but

the vulnerabilities and inequalities it has revealed and widened will remain—as will the need for social protection to be adapted for and better targeted to those affected by the inevitable shocks of the future, which climate change will only heighten in frequency and intensity.

Questions about whether and how to target were pressing before COVID-19 and will remain so long after its immediate impacts recede. New technology and data have changed much in the modern world and are an important development that receives considerable treatment within this book. New policy practices and new papers were emerging monthly as the book was written, and the examples discussed are likely to be soon superseded. Nevertheless, it is hoped that the questions the book raises and the principles it outlines will remain important and germane for many years.

Michal Rutkowski
Global Director, Social Protection and Jobs Global Practice
The World Bank

Acknowledgments

About the Editors and Contributors

Editors

Margaret Grosh is the Senior Advisor for the World Bank's Social Protection and Jobs Global Practice. She has written, lectured, and advised extensively on social protection programs, especially on targeting and cash transfer programs, globally and for Latin America. She has extensive experience with social protection both for responding to a crisis and for improving equality of opportunity. Prior to her current position, she served as Practice Manager for Social Protection and prior to that as Lead Economist in the World Bank's Latin America and Caribbean Region Human Development Department. Earlier, she led the Social Assistance team in the World Bank's Global Social Protection Department and the Living Standards Measurement Study in its Research Department. She holds a PhD in economics from Cornell University.

Phillippe Leite is a Senior Social Protection Economist in the South Asia Unit of the Social Protection and Jobs Global Practice at the World Bank. He has worked extensively on support design and implementation of social assistance programs, social registries, and delivery systems in Africa and Latin America. Before joining this Global Practice, he worked in the World Bank's Development Research Group on determinants of poverty and inequality, methodology for poverty maps, and microeconometric simulation models. He holds a BA and an MS in statistics (Population Studies and Social Research) from the Escola Nacional de Ciências Estatísticas-ENCE (Rio de Janeiro), and a Diplôme d'Études Approfondies and a PhD in

economics from DELTA-École des Hautes Études en Sciences Sociales, currently the Paris School of Economics (Paris).

Emil Tesliuc is a Senior Economist with the Global Engagement unit of the Social Protection and Jobs Global Practice at the World Bank. He has worked broadly across social protection on the design and implementation of social assistance programs, social care services, employment support programs, digital service delivery, reducing error and fraud in social protection programs, and parametric social protection reforms. He has experience in Europe and Central Asia, the Middle East and North Africa, and Sub-Saharan Africa. He currently leads the *Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE)*, a global database that monitors the size and distributional performance of social protection and labor programs in World Bank client countries. He holds a PhD in economics from the Bucharest University of Economic Studies and an MA in public policy from Princeton University.

Matthew Wai-Poi is a Lead Economist with the Poverty and Equity Global Practice at the World Bank and has worked in East Asis and the Pacific and, the Middle East on poverty, inequality, and their determinants; the middle class and top incomes; gender; forced displacement; and climate change. He is also Global Lead for Distributional Impacts of Fiscal and Social Policies and has authored various reports and papers on targeting and supported the development of a number of national targeting systems. He holds a PhD in economics from Columbia University and degrees in law and business, and he previously worked in management consulting.

Contributors

Ana Areias is a Development Economist and Data Scientist working at the intersection of machine learning and public policy. She has collaborated with the World Bank to apply new methodologies from machine learning in settings such as survey-to-survey imputation for the Poverty and Equity Practice and the development of proxy means testing tools for Liberia with the Social Protection and Jobs Global Practice. She worked with the Education Global Practice using big data from an Indian job-matching website to understand wage determinants and the gender wage gap and to perform a randomized control trial to increase the efficiency of job matching. Currently a data scientist at Kineviz, she works on merging deep learning and graph technologies to better understand high-dimensional and interconnected data. She has also been a Fellow at Data Incubator, a member of DataCorps with DataKind, and Program Manager at Data-Pop Alliance. She holds an MPA in international development from Harvard University.

Priyanka Kanth is an Economist who has worked across the social protection and health sectors at the World Bank. She has experience working in Africa, Middle East and North Africa, and South Asia. In social protection, she has worked on supporting the design and implementation of safety net programs and social registries and carried out analytical work on labor and skills, savings and risk mitigation, and impacts of different policy initiatives on human capital indicators more broadly. She has worked on health service delivery and health financing issues. Before joining the World Bank, she completed research assignments at Stanford University and the Jameel Poverty Action Lab. She has master's degrees in health policy jointly from the London School of Economics and the London School of Hygiene and Tropical Medicine, as well as in development economics from Yale University, where she was a Fulbright Scholar and is completing her PhD in economics.

Juul Pinxten is a Social Protection Specialist in the Middle East and North Africa unit of the Social Protection and Jobs Global Practice at the World Bank. He supports social assistance policy and delivery system reform advocacy in the Arab Republic of Egypt and the Republic of Yemen. Prior to joining this unit, he led a social protection engagement in Timor-Leste and worked on informing the design and implementation of cash transfers, food assistance programs, and the social registry in Indonesia. He holds a BA in liberal arts from Maastricht University and an MS in public policy and human development from the Maastricht Graduate School of Governance and United Nations University.

Claudia P. Rodriguez Alas is a Social Protection Specialist in the World Bank's Social Protection and Jobs Global Practice. Her work at the World Bank focuses on generating global knowledge products on social protection and leading the *Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE)* project. She previously worked at the Superintendency of Pensions of El Salvador, where she drafted regulations for the newly reformed social security system. She has also worked with nonprofit organizations on community outreach and immigrants' rights in the Washington, DC, metropolitan area. She received her bachelor's degree in economics from Montana State University, where she was a Fulbright Scholar. She also holds a master's degree in international development from American University in Washington, DC.

Nina Rosas is a Senior Economist in the South Asia unit of the Social Protection and Jobs Global Practice at the World Bank. She currently focuses on the World Bank's operations in Pakistan, supporting the government in the implementation of large social assistance programs and introduction of hybrid social insurance mechanisms. Prior to this, she supported the design,

implementation, and evaluation of social assistance, youth employment, and delivery system interventions in West and East Africa. During this period, she led the World Bank's emergency response to the Ebola crisis in Sierra Leone from the social assistance side. She was a team member of the World Development Report 2012 on gender and development and before joining the World Bank, she worked in PricewaterhouseCoopers' Global Restructuring Group. She holds a bachelor's degree in economics from the University of San Diego and a master's degree in public administration/ international development from Harvard University's Kennedy School of Government.

Usama Zafar has been working with the Social Protection and Jobs Global Practice at the World Bank since 2018, supporting the *Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE)* project on its data enhancements and enrichment efforts and helping multiregional teams with their analytical work. He has also worked with the United Nations Development Programme to help assess the socioeconomic impact of the COVID-19 crisis in the Pacific Islands. Prior to joining the World Bank, he was involved with a political consulting group in Washington, DC, where he used his passion for data analytics to design campaign strategies for political candidates. He holds a master's degree in public policy from Texas A&M University and a bachelor's degree in economics from Lahore University of Management Sciences in Pakistan.

Abbreviations

4Ps	Pantawid Pamilyang Pilipino Program (conditional cash transfer program) (Philippines)
ANPIS/ANPOFM	Ministry of Labor and Social Policy/National Agency for Payments and Social Inspection (Romania)
APEC	Asia-Pacific Economic Cooperation
ASPIRE	Atlas of Social Protection: Indicators of Resilience and Equity
BISP	Benazir Income Support Programme (Pakistan)
BPC	Benefício de Prestação Continuada (Brazil)
BPNT	Bantuan Pangan Non-Tunai (digital food voucher Food Assistance Program) (Indonesia)
CadÚnico	Unified Registry of Social Programs (Brazil)
CAGED	Cadastro Geral de Empregados e Desempregados (Brazil)
CAS	Circonscription d'Action Sociale (Local Social Assistance Offices) (Republic of Congo)
CBT	community-based targeting
CDR	call detail record
CEGA	Center for Effective Global Action (University of California, Berkeley)
CGD	Center for Global Development
CMP	Child Money Program (Mongolia)
CONPES	National Council for Economic and Social Policy (Colombia)

CRES	Centre des Recherches et des Études Sociales
DCI	Distribution Characteristic Index
DFAT	Department for Foreign Affairs and Trade (Australia)
DIME	Development Impact Evaluation
DNP	National Planning Department (Colombia)
DSD	Department of Social Development (South Africa)
ECLAC	Economic Commission for Latin America and the Caribbean
eCNY	electronic Chinese yuan
ECT	emergency cash transfer
EHCSN	Employment and Human Capital Safety Net (Djibouti)
eID	electronic identification
ENIGH	National Survey of Household Income and Expenditure (Mexico)
EU	European Union
F_1	performance measure that weights inclusion and exclusion errors equally
F_2	performance measure that weights exclusion errors twice as much as inclusion errors
FCDO	Foreign, Commonwealth, and Development Office (United Kingdom)
FGT0	Foster-Greer-Thorbecke Poverty Headcount
FGT1	Foster-Greer-Thorbecke Poverty Gap
FGT2	Squared Poverty Gap
FIBE	Emergency Basic Fact Sheet (Chile)
Ficha CAS	eligibility form for poverty-targeted social assistance programs (Chile)
GAO	Government Accountability Office (United States)
GDP	gross domestic product
GRM	grievance and redress mechanism
GSMA	Global System for Mobile Communication Association
HEA	Household Economic Analysis
HMT	hybrid means test
ID	identification
ID4D	Identification for Development
IDA	International Development Association
ILO	International Labour Organization
IMF	International Monetary Fund
ISAS	Integrated Social Assistance System (Turkey)
ISSA	International Social Security Association
IT	information technology
ITU	International Telecommunication Union
KPH	Family Hope Program (Indonesia)

LOOCV	leave-one-out cross validation
LPG	liquefied petroleum gas
LSMS	Living Standards Measurement Study
MCC	Matthews correlation coefficient
MINALOC	Ministry of Local Government (Rwanda)
MISSOC	Mutual Information System on Social Protection (European Commission)
MoARD	Ministry of Agriculture and Rural Development (Ethiopia)
MoLSA	Ministry of Labour and Social Affairs (Armenia)
MPI	Multidimensional Poverty Index
NASSCO	National Social Safety Net Coordinating Office (Nigeria)
Novissi	Emergency Cash Assistance Program (Togo)
NRMSE	normalized root mean square error
NSER	National Socio-Economic Registry (Pakistan)
ODI	Overseas Development Institute
OECD	Organisation for Economic Co-operation and Development
OLS	ordinary least squares
OPM	Oxford Policy Management
PATH	Program of Advancement Through Health and Education (Jamaica)
PIS	Social Integration Program (Brazil)
PMT	proxy means test
PNAD	National Household Sample Survey (Brazil)
PNSF	Programme National de Solidarité Famille (National Programme of Family Solidarity) (Djibouti)
PSNP	Productive Safety Net Program (Ethiopia)
RAIS	Relação Anual de Informações Sociais (Brazil)
RAMED	Medical Assistance Plan (Morocco)
RMSE	root mean square error
RSH	Registro de Social Hogares (Integrated Social Information System) (Chile)
SAAIS	Social Assistance Automated Information System (Moldova)
SASSA	South African Social Security
SD	standard deviation
SDG	Sustainable Development Goal
SISBEN	System for the Selection of Beneficiaries for Social Programs (Colombia)
SIUBEN	Social Registry (Dominican Republic)
SNAP	Supplemental Nutrition Assistance Program (United States)

SPACE	Social Protection Approaches to COVID-19: Expert Advice
SPIAC	Social Protection Inter-Agency Cooperation
SSA	Social Security Allowances (Nepal)
SUF	Subsidio Único Familiar (Chile)
SWC	Social Welfare Center (Croatia)
TANF	Temporary Assistance to Needy Families (United States)
TD	targeting differential
TSA	Targeted Social Assistance Program (Georgia)
UNECA	United Nations Economic Commission for Africa
UNICEF	United Nations Children's Fund
UNRISD	United Nations Research Institute for Social Development
USP	universal social protection
UTGFS	Unité Technique de Gestion Filets Sociaux (Mali)
WHO	World Health Organization
WIC	Women, Infants, and Children (United States)

Overview

Margaret Grosh, Phillippe Leite,
Matthew Wai-Poi, and Emil Tesliuc

PREMISE: Most countries target some social protection programs to selected people; reviewing the current knowledge on this subject can inform the formulation of policy

Social protection systems, the social contracts of which they are a part, and the mix of institutions and programs used to achieve their goals vary greatly from country to country. Within that diversity, three constants shape the discourse and practice around prioritizing those in need.

First, there is a strong consensus around the need to reduce poverty and inequality and a drive toward universal social protection (USP). That consensus is reflected in many national policy statements and even some constitutions mandating USP, and in the rapid expansion of social protection programming and many innovations seen in recent years. The goal of USP has been codified as part of the Sustainable Development Goals to be met by 2030 and supported by a long list of international and bilateral organizations, including the World Bank, which are partners in the USP 2030 initiative.[1]

Second, it is a fact of life that hundreds of social programs around the world differentiate eligibility and/or benefits in various ways. Many countries have multiple programs that base eligibility or differentiate benefits according to welfare levels, and often one or more of these are high-profile flagship programs. Nearly every country has at least one poverty-targeted social assistance program. Many countries have special programs to support children and the elderly, because they are deemed biologically or socially vulnerable,

1

more likely to be poor, or both. The unemployed may benefit from unemployment insurance, and those struck by natural disaster may receive assistance initially to sustain them and eventually to help them rebuild housing or livelihoods. Productive inclusion programs seek to raise the level or decrease the variability of the incomes of the poor. Active labor market policies usually focus on those with greater barriers to (re-)employment. Thinking more broadly about people lagging in education or without access to health care or essential utility services, many more programs direct various social policy efforts to members of these groups to improve their life outcomes. These targeted programs assist in achieving the goals of universal coverage and sit next to universal programs in broader social policy, with the mix of universal and targeted programs varying from country to country.

Third, countries choose to differentiate eligibility and benefits to fit programs to purpose and reduce their costs, but the job of targeting individuals or groups is fraught with conceptual and practical difficulties, it has errors and costs, and there are many criteria and metrics with which success or lack thereof can be gauged. Thus, the issue of whether current practice is acceptable, can be improved, or should be abandoned recurs in instance after instance.

The tension between and within these three constants makes the choices around whether and how to target those in need of different facets of social protection a perennial topic in policy discussions and is the motivation for this book. Because the role, how-tos, and difficulties in targeting are recurring features of social protection practice and debate, it is important to learn as much as possible so that policy choices can be well informed. Responses to the tension have varied from place to place and over time depending on each country's challenges and resources and the specific progress and gaps in its social protection system to date. This has generated a rich set of global experience from which countries can draw as they review and renew their progress toward USP.

The basic dilemmas and general findings around targeting are familiar from earlier literature. Targeting is a tool, not an objective. Its use must be judicious—weighing the potential gains, errors, and costs. Different targeting methods have different potentials for accuracy and costs. The magnitudes of these are context specific, so decisions that are appropriate in one setting may not be in another. Perfect (errorless) targeting is more a useful abstraction than an empirically observed benchmark, and certainly in the face of costs, it is not optimal or necessarily preferred. Moreover, targeting is only one of several important design parameters. The budget greatly influences the potential of the program; the benefit or quality of the service is also important for achieving the objectives of a program.

It is worth reexamining issues of targeting as both social protection practice and technology have evolved. The social protection community has

new information to illuminate trade-offs, new implementation capacities, and the potential to bring new data and data science to bear, so it is useful to update the treatment of old issues. The goal of this book is to do so.

The aim of the book is to inspire more success where the choice is made to target. It is hoped that the book will help the reader understand the many facets of the choices, become familiar with global experience, be able to discern how well different options might work in a particular context, and learn how to customize approaches to context. The variety of practices and some clear improvements in aspects large and small where thoughtful or diligent effort has been made suggest that there is room for greater successes in many places. Capacities built and the secular trends in data and technology should make it possible to raise the game.

It is not the aim to push the reader toward one degree of targeting versus another or one method versus another. The book aims to equip the reader to discern well what is appropriate in the context and for the program on which she or he is working. Appropriate choices depend on the purpose, context, and capacities, and these vary widely across the range of programs encompassed in social protection and the diversity of countries.

The book makes a technical contribution to targeting, which is only one design element of a program and is used to different degrees in different places for different programs. A full consideration of how to design well each kind of program (for example, whether it should be in cash or in kind; the level of benefits; whether to have any nudges, conditions, or linked services; and so forth) or customize those answers to particular settings is not given. The book does not speak to the full social protection policy package, which encompasses a variety of programs to support people throughout the life cycle, across the welfare spectrum, to address a range of vulnerabilities and risks and help countries achieve resilience, equity, and opportunity (for the World Bank's social protection strategy, see World Bank 2012, 2021).

The book focuses on programs that intend to differentiate eligibility or benefit levels along the spectrum of money-metric welfare and the shocks that affect it. It looks at methods that measure or infer these household by household and those that use more aggregated geographic or demographic categories that may be correlated with money-metric poverty or vulnerability. The idea of targeting applies more broadly—to determining who is more in need of policy supports in some way, to include those living with a disability, those lagging in employment or education, or those without access to health care or essential utility services. Although there are several parallels or analogues between money-metric and other types of targeting—especially in thinking about tolerance for inclusion or exclusion errors, the need to build good administrative systems and concern about their cost, the need to consider the human rights and behaviors of those involved, and so forth—the tools of discernment between poverty, disability, employability,

and other factors are different. Thus, to keep the topic manageable, the book focuses on those programs related to money-metric welfare.

Drawing together literature from poverty diagnostics, disaster risk management, social assistance practice, and evaluation, the book includes references to tax policy, political economy, and human rights, which shape choices around how much and how to target. Based on a great deal of literature that is rigorous in its statistical methods, the book is largely written in a nontechnical manner meant to be readable by people from the several disciplines that think about social protection policy. The last three chapters contain more technical material, with the heaviest confined to the latter parts of the chapters and/or appendixes. Nonstatisticians should be able to follow the highlights of all the chapters, perhaps skipping or skimming some of the deeper dives into more technical material.

The book is organized as follows. This chapter provides a brief introduction to the framing and terminology and an overview of the messages emerging from the book as a whole. Chapter 1 presents a series of essays on the factors that shape the choices around why, whether, or how narrowly/broadly to target different parts of social assistance. Chapter 2 updates the global empirics around the outcomes and costs of targeting benefits on the poor or vulnerable. Chapter 3 illustrates the options and choices that must be made in moving from an abstract vision of focusing resources on the poor or vulnerable to more specific concepts and implementable definitions and procedures, and how the many choices should be informed by values, empirics, and context. Chapter 4 provides a brief treatment of delivery systems and processes (more traditionally known as implementation or administrative systems), showing their importance for targeting outcomes and suggesting the many frontiers with room for improvement. Chapter 5 discusses the choice among targeting methods and how contexts and factors shape them. Chapter 6 summarizes and comprehensively updates the know-how on the data and inference used by the different household-specific targeting methods. Chapter 7 contains a primer on measurement issues, going much deeper than usual and explaining how better measurement can lead to clearer understanding of targeting issues. Chapter 8 explores machine learning algorithms for household-specific mechanisms for eligibility determination.

Framing and Terminology: From Objectives to Outcomes

To clarify concepts, the book uses a hierarchy of policy objectives, program design, targeting methodology, implementation, and metrics for outcomes. A policy objective is an overarching goal, such as "education for all" or "reduction of poverty." A program is an intervention that is implemented to

achieve the policy objective and can be population wide or for a subset of the population. The implementation consists broadly of all the other components that make a program work. A targeting methodology is a tool to identify the intended population for a certain program, to conduct eligibility assessments. The targeting metrics measure how well the program reached the intended population and the distributional outcomes. Evaluation metrics help to discern how the program changed other key outcomes. Figure O.1 provides an overview of these elements. Box O.1 is a note on terminology.

Figure O.1   Hierarchy of Action

EXAMPLES

Policy objectives
- Poverty reduction
- Education for all
- Female empowerment
- Universal social protection

Program level
- Provide a specific set of services or benefits to a defined intended population, for example, income support to the poor, scholarships to out-of-school youth, or reproductive health services to adolescents
- Generally, multiple programs help to achieve a single goal
- Similarly, single programs may contribute to multiple goals

Delivery system
- Outreach
- Assessment of needs and conditions
- Enrollment
- Provision of benefits or services

Targeting method
- Self-targeting
- Geographic
- Demographic
- Welfare based
  - Means test
  - Hybrid means test
  - Proxy means test
  - Community-based targeting

Targeting metrics
- Dichotomous, for example, errors of inclusion, errors of exclusion, and targeting differential
- Continuous, for example, coverage, incidence, and distributional characteristic
- Costs, for example, administrative costs, transaction costs, stigma or social discord, and incentives

Impacts
- On poverty and distribution
- On human capital
- On labor activity

Source: Original compilation for this publication.

BOX O.1

Terminology

The topic of differentiating eligibility and benefits by some dimension of need is commonly referred to in social protection jargon as "target-ing." In English, the term "targeting" may sound a little fierce since the same term is used in social policy, advertising, and shooting. Some languages escape such unhappy association. In Polish, Ukrainian, and Russian, the terms *adresowanie* (Polish), *адресність* (Ukrainian), and *адресность* (Russian) are used, which are cognates of the term used for postal addresses. That is more neutral on the emotional register and aptly implies specificity to an individual or family. In Spanish, *focalizar* and *focalizacion* are neologisms that do not exactly trip off the tongue, but with their root in "focus," as in concentrating benefits on needy populations, they also have a more positive connotation and come quite close to the economic rationale for why to target.

To reduce the negative connotations in English, this book somewhat reduces the use of the word targeting, using substitutes such as the differentiation of eligibility or benefits. The term targeting is not eschewed altogether as it is a well-known and sometimes handy short-hand for the end-to-end process, embedded in the names of different methods and metrics.

Changing the term does not reduce the complexity of the topic or the difficulty of making judgments or gaining consensus. The choices to be made are consequential, the processes are subject to many trade-offs and can be judged by different metrics, and so the topic remains subject to contestation.

The book uses the term "welfare" often as shorthand, to refer to one or more of a suite of related but different specific definitions. Chapter 1 uses it as shorthand for the general concept of monetary well-being. Chapter 2 also refers to monetary welfare, in this case, as recorded in the household surveys in the Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE) database, measured as income in some countries and consumption in others. Chapter 3 takes up more directly the different dimensions of monetary and nonmonetary welfare. The later chapters revert back to a narrower monetary concept of targeting. Chapter 6 examines in a bit more detail the options and implications of how to define and measure money-metric welfare as a method of eligibility determination.

Social protection programs often have multiple outcomes and can cater to multiple policy objectives; similarly, a policy objective could be supported by several different programs. For instance, a conditional cash transfer program could cater to the objectives of improving both education outcomes and household consumption for the poor.

USP is an overarching policy objective that can be met through layering different programs that address different population needs. It could be achieved through a combination of universal health care, income support for the poorest, old-age pensions, unemployment benefits, skills development for the unskilled, and more. Some of the programs and policies have redistribution objectives and these need to be focused on those who are in the lower end of the welfare distribution. Some programs with risk management or productivity goals affect those throughout the welfare distribution, are geared to those with the highest risks, or focus on those for whom risk has become manifest in a negative shock.

To discern how to achieve improvements, it is important to understand where there may be missteps along the chain from the desired objective to the actual outcome. For example, a program could cover a small share of the poor. To understand to what degree that is a good or bad outcome or how to improve it requires considering the program's objective, size, implementation, and benefit level, as well as the constraints that shape these. Was it meant to cover all the poor? Or only a subset, for example, poor widows? Is the program funded at a level sufficient to cover all the poor? Does the program include requirements that the poor are unlikely to meet? Is the delivery system more inclusive or less so? Is the eligibility determination process reliable? Although the focus of this book is on the middle and later links in the hierarchy of action, it tries to be clear about the dependence on prior steps.

Synopsis of 10 Key Messages

Message 1. Targeting selected categories, families, or individuals can play a valuable role within the framework of USP.

The most commonly used rationale for targeting is that for a given budget, a larger impact can be achieved when more resources are focused on the needier. The storyline is almost arithmetic, and it is most commonly told for poverty: the basic formulation of x million focused on the poor will lower poverty more than if the same budget were spread among more people or universally. Analogously, the impacts on

inequality will be greater the more concentrated are the resources at the lower end of the income distribution. Moreover, theory and evidence support the idea that the impacts of social assistance transfers on various dimensions of human capital and productive activities may be greater for the poorest as well. A similar but less commonly articulated reason for targeting is to make programs for risk management fit for purpose. For such programs, the focus must be on those at greatest risk of adverse events and/or bad outcomes from them. Of course, knowing who falls within the group of highest priority is not easy, generating both errors and costs. Thus, the theoretical promise of targeting can only be partially realized. This basic issue of how intensely to concentrate or how widely to allocate budgets is thoroughly embedded in discussions of social protection, including the feedback loop of policy choice, political support, and the budget. Chapter 1 contains the most focused discussion of how targeted programs fit into USP. The country and program examples throughout the rest of the book provide ample illustrations.

As countries strive toward USP, the weight given to programs that target benefits on particular groups can be quite varied. It depends on the development of other programs, degree of poverty and inequality, sources and magnitudes of vulnerabilities and risks, level and structure of revenues, political economy, and history. In countries with high levels of employment generating good earnings, high coverage of social insurance, and perhaps full coverage of children and the elderly with allowances and pensions, many people are protected from poverty and risk by their earnings and insurance. In such countries, the share of social protection expenditures on poverty-focused or last resort social assistance may be low, but it is still important to assist those who fall between the cracks of other social efforts and complete the universality of the guarantee of assistance. In other countries, many workers may remain in poverty or in a vulnerable situation due to low productivity in their employment and are not covered by social insurances. In some countries, many people cannot benefit from existing policies and are constantly affected by the recurrent nature of some shocks, which can push people (further) into poverty. In such cases, it may be best for the initial development of the social protection sector to start by focusing on the poorer and more vulnerable to shocks before expanding to others.

Both the level of taxation and its incidence should be taken into account in considering how extensive or redistributive transfers may be. It is net positive transfers that may reduce poverty and monetary inequality or help reduce the gradient in human capital outcomes and vulnerabilities across income levels, so the less progressive are a country's tax systems, the more important it will be to have some progressivity on the

expenditure side. Countries that generate more revenue can afford more complete social protection systems and have more leeway to choose universally or affluence tested designs over more narrowly targeted ones for a wider range of their programming. In countries with particularly low expenditure on social protection, there is a particular need to focus on raising resources; even the power of targeting cannot make up for gross insufficiency of funding.

The human rights framework prods countries to establish more substantial social protection and pay more attention to high-quality implementation of targeted programs. The human rights framework demands that governments make maximal efforts to ensure sufficient budget and programming to realize the social and economic rights of people living in the country. Thus, the framework takes a somewhat different vein from the economically stated problem of the best use of a fixed budget. The human rights framework does not preclude differentiation of eligibility per se, noting the need to focus first on the especially disadvantaged as part of progressive universalism. However, the framework sets high standards for the implementation of such mechanisms, standards that although hard to meet are desirable from other points of view as well. The human rights framework puts very heavy weight on eliminating errors of exclusion and ensuring accessibility, inclusion, nondiscrimination, dignity, gender sensitivity, privacy, transparency, and accountability in the implementation of policy. Although few programs can claim an excellent record on all these criteria, they may be easier to meet for some targeting methods than others or for some criteria with one method and for other criteria with another. Improvements in delivery systems are critical for improving the respect for human rights within programs that differentiate eligibility by need. Indeed, the topic of human rights and differentiated eligibility and benefits are intertwined, so throughout the book, discussion of human rights is interwoven within the economic, statistical, and institutional treatments.

The coronavirus (COVID-19)-induced economic shock brought the question of the scope and shape of social protection to higher political visibility across more countries simultaneously than perhaps has ever before happened. In response to the economic damage that followed in the track of the novel coronavirus, nearly all countries took multiple actions to expand their social protection systems in coverage, benefit levels, or both, sometimes raising them by substantial amounts. Although many responses were temporary, by reaching so many people so quickly, the responses may have heightened expectations for social protection going forward. Countries face periods of recovery, of as yet unknown and likely variable speed and duration. Many also face increased deficits and debt loads, so budget space to support larger social protection and meet the heightened expectations

may not be easy to find. Thus, the desirable degree of targeting, as a way of mediating between great needs and scarce resources, may remain a prominently debated issue. Indeed, the work of building social protection systems to reach more people and customize support to them as their circumstances differ among each other and over time may get more attention than it ever has in the past.

Message 2. Measuring the accuracy and costs of targeting can be done in many ways; judicious choices will consider a range of them.

The accuracy of eligibility determination processes has many facets, which are captured by a range of metrics, and there is some lack of consensus within the social protection community about which outcomes are most important. Quantitative metrics focus on who would benefit and who would not, how it would change their welfare, and the costs of transactions for beneficiaries or program administrators. More qualitative metrics look at political economy, social cohesion, and stigma.

A divisive issue in many discussions around targeting is how much weight to give to errors of exclusion (those who do not receive benefits but are part of the intended population) versus errors of inclusion (those who receive benefits but were not in the intended population). Consider two thought experiments:

- *With a smaller budget than what is needed to cover all the poor*. Imagine a scenario that starts with a baseline of whatever poverty levels there were before the initiation of a possible new social assistance program, which will at least initially be budget rationed to include fewer than all the poor in the country. In such a scenario, a poor person who is excluded due to a mistake in eligibility assessment ends up with the same welfare[2] as all those who are unserved because the program budget is insufficient to cover all the poor. A needy person served by the program is better off. Any nonpoor person included (error of inclusion) takes up budget that could have helped a poor person. In this scenario of budget rationing, reducing errors of inclusion is a means to reduce errors of exclusion, and errors of exclusion are inevitable given the rationing. This scenario fits well the situations in many poor countries just starting to build their social protection systems. In such cases of budget rationing, including nonpoor people in the programs will crowd out the poor. To cover all the poor, increasing budgets is vital, but reducing errors of inclusion will help as well.
- *With enough budget to cover the poor and the unavoidable inclusion errors.* This scenario starts with a budget that is sufficient at least to serve all those who are poor plus any nonpoor people who are in the program by design

or due to errors in eligibility assessment. In this case, reducing errors of exclusion is vital to ending poverty and realizing the principle of nondiscrimination as articulated in the human rights frameworks, but the ability to do so is not rationed by the budget, only by potential deficiencies in the delivery system or eligibility determination mechanism. Reducing errors of inclusion may reduce costs or improve the program's reputation, but with a budget already sufficient to serve all the poor, it will not map directly into reducing errors of exclusion.

Instead of focusing only on errors of exclusion or inclusion, it is important to consider a fuller distributional analysis in evaluating the outcomes of social assistance programming, as well as to consider program size and design. It is preferable to use measures that consider the full distribution of the program's benefits. This makes it possible to give more weight to an error of exclusion at, say, the 5th centile than to errors that might occur around a threshold, say errors around the 19th centile in a program meant to reach 20 percent of the population. It also makes it possible to weigh with more tolerance an error of inclusion at, say, the 25th centile than the 75th centile. Moreover, examining the full distribution makes it possible to contrast programs with different funding levels and eligibility thresholds.

A common way to make judgments about different kinds of errors is to base the judgments on how different measures of poverty—such as the headcount or poverty gap—are affected. If a few of the intended population are excluded but enough extra value can be given to the poor(er) who are included, it will reduce overall poverty or inequality and thus be judged better for society as a whole. As illustrated in chapter 5, it is recommended to use at least the poverty gap rather than only the headcount as the poverty gap gives greater weight to the poorer and values increases in welfare even among those who may not be boosted across the poverty line by a program's benefit. More complete measures, such as the distribution characteristic, which considers the whole welfare distribution rather than just those below the poverty line/eligibility threshold, are well codified in the welfare economics literature, although they are less frequently used in policy dialogue (see chapter 7). Different statistical modeling techniques to inform proxy means testing can vary the weights given to observations in different parts of the income distribution or to errors of exclusion versus errors of inclusion (see chapters 6, 7, and 8).

Judgments about empirical findings must account for any limitations on what can be observed in the evaluation data being used. These are often estimated from general purpose household surveys and may differ from a program's eligibility determination process in the definitions of the unit of observation, measures of well-being, timing of the observation, and so forth. It is also important to understand the program context, rules, and implementation procedures to draw appropriately nuanced conclusions.

Message 3. Surveying international targeting outcomes shows that social assistance coverage is incomplete but progressive, although there is wide variation among countries and programs. To reduce poverty, it is usually more cost-effective to ensure that a greater share of benefits accrue to the poor than to expand coverage broadly.

To help illuminate the choices around targeting, it is important to understand not just the theory, but also the empirics of the trade-offs involved. The many different parts of the whole empirical story are spread across the chapters of the book.

Chapter 2 reports on a broad overview of the coverage, incidence, and simple[3] estimates of impacts on poverty of a wide range of social assistance programming in developing countries, using the Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE) global data set.[4] The most recent data (from 2014–19) are used, covering 70 World Bank client countries. This data presentation casts the net widely to look at emerging and developing countries with recent household survey data and all kinds of social assistance programs, irrespective of the intention to target benefits, the eligibility thresholds, or the targeting methods used. The analysis provides benchmarks for the outcomes observed, which can be useful comparators for country-specific discussions and setting expectations about the feasibility of different scenarios. For selected observations, the chapter supplements the broad picture that is observable from the surveys with a bit of program detail to contrast some of the choices and outcomes, foreshadowing the deeper discussion on such choices in subsequent chapters.

Social assistance coverage among the observed developing countries is incomplete but concentrated on the poor: 36 percent of the overall population receives a social assistance benefit of some kind, whereas 54 percent of people in the poorest quintile do. Coverage generally increases as country income does but with high dispersion in this pattern. In low-income countries, on average, only 17 percent of people in the poorest quintile are covered by social assistance, whereas the figure is 77 percent for high-income countries.[5] The range of coverage overall is broad, from virtually none to virtually all of the poorest quintile covered. There are truncations of the range at the extremes of the country income groups, none of the low-income countries manages to cover more than about half of the poorest quintile, and none of the World Bank client high-income countries covers less than half of the poorest quintile.

There is a great deal of variation in the share of benefits accruing to the poorest quintile, irrespective of which program type is considered. Incidence graphs aggregated across countries by program type look more like a mildly downward-sloped hill than a sharp step function. In these program-type aggregations, the share of beneficiaries in the poorest quintile ranges from

30 to about 50 percent depending on program type. There is even more variability at the level of the individual country and program type. For unconditional cash transfers, the range is from 6 to 73 percent of the beneficiaries being in the poorest quintile of the population. Analogous outcomes for other program types are discussed in chapter 2. They too show large variation, although the range and number of observations are largest for unconditional cash transfers. This may reflect the wide variety of programming that falls under this category, as well as the varied degrees to which they are captured in the surveys.

Greater coverage of poor households can be driven by concentrating benefits on the poor, increasing the size of programs, or both. The previous paragraphs highlighted the wide variation in coverage of the total population, coverage of poor households, and share of the benefits that goes to poor households. We now consider how these three factors interact. Figure O.2 plots the coverage of the poorest quintile (represented as the size of the bubble) as programs increase in total coverage (along the x-axis) and increase in share of benefits concentrated on the poorest quintile (along the y-axis). The first notable feature is that the bubbles tend to get larger moving from left to right; that is, larger programs generally cover more of the poor. The second feature of significance is that the bubbles get larger moving from the bottom to the top; that is, more progressive programs cover more poor households for a given program size. A feature in figure O.2 is that the bubbles are widely dispersed from top to bottom and become compressed among the range of smaller programs. If a program is smaller than the population it is ideally meant to serve (here defined as the bottom quintile), it will not cover enough of the poor even when they receive a large share of the benefits. The bubbles for programs that cover less than 10 percent of all people are all relatively small, even when the share accruing to the poorest quintile is high. That is, although targeting can help smaller programs reach more of the poor, exclusion errors will remain high due to the program size. If a program is very large, it may cover more of the poor, but also many of the nonpoor; at most, a third of the benefits goes to the poorest quintile for the programs covering more than half of the population.[6] In the medium-size programs, there is less predetermination about whether errors of exclusion or inclusion will dominate.

Concentrating benefits on the poor is a more cost-effective way of reducing poverty than simply increasing program size. To understand the relative roles of incidence and program size, figure O.3 overlays on a single grid the share of beneficiaries in the poorest quintile and the total program coverage of the population, both plotted against the poverty gap reduction per $1 spent. Thus, each program observation appears on the graph twice, in different colors. For example, in the upper left and upper right in the figure, with a value of 0.8 on the y-axis, two dots are labeled PAN. These represent the

Figure O.2 Incidence and Coverage of Cash Program Beneficiaries, as Captured in the ASPIRE Household Surveys, Including Very Large Programs

Source: ASPIRE database, www.worldbank.org/aspire.

Notes: The number of countries per region with cash programs is as follows: world (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 49), conditional cash transfers (n = 20), social pensions (n = 38), public works (n = 8). Coverage is calculated as (number of individuals in a given group [for example, total population or poorest quintile] who live in a household in which at least one member receives the transfer)/(number of individuals in the group). Beneficiary incidence is calculated as (number of direct and indirect beneficiaries [people who live in a household in which at least one member receives the transfer] in a given quintile)/(total number of direct and indirect beneficiaries). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption).

Panamanian conditional cash transfer program category, which is dominated by the Red de Oportunidades program, a smallish but fairly traditional Latin American conditional cash transfer program.[7] The 0.8 means that the poverty gap is reduced by 80 cents for each dollar spent on the program. The dark blue dot (coordinate 82.7, 0.8) shows that about 83 percent of the beneficiaries are in the poorest quintile; thus, it is the most progressive

Figure O.3 Efficiency of Cash Programs, as Captured in the ASPIRE Household Surveys

Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region with cash programs is as follows: World (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 43), conditional cash transfers (n = 15), social pensions (n = 35), public works (n = 6). Monetary information was not available for 16 programs to generate poverty gap indicators. For this reason, the sample of cash programs used for this figure is smaller than the one used for figures estimating coverage and incidence of benefits. Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption). PAN = Panamanian conditional cash transfer.

observation. However, the program's total coverage is low; the light blue dot (coordinate 6.6, 0.8) shows that it covers just under 7 percent of the total population (and therefore just 27 percent of the poorest quintile). And so on for all the observations or programs/program types and countries.

There is a much stronger relationship between the incidence of benefits and the reduction in the poverty gap per dollar spent than between overall coverage of the population and the reduction in the poverty gap per dollar spent. In figure O.3, the dark blue line, representing the incidence of benefits for the poorest quintile, increases relatively steeply, indicating that as a greater share of program benefits goes to the poorest quintile, the degree of reduction in the poverty gap for each dollar spent increases; higher incidence gives a bigger bang for the program buck. Specifically, an additional 10 percent share going to the poorest quintile means that the poverty gap falls by an extra $0.10 per $1 spent. Conversely, the light blue line, representing total program coverage of the population, rises only modestly until programs reach around 30–40 percent population coverage and then declines. Although there is very high variation among smaller programs, ranging from best to worst performers in terms of efficiency of poverty reduction, the largest programs reduce the poverty gap the least per dollar spent; at some stage, they run out of poor people to cover.

Message 4. The various costs of targeting selected groups, families, or individuals are usually low or within an acceptable range.

It is well accepted that targeting has costs and these must be balanced against the potential gains of focusing resources on those most in need. Perhaps the most pervasive concern is the political economy, but labor disincentives and administrative costs are often prominent concerns as well. Stigma, social cohesion, and transaction costs tend to receive less attention but are of course broadly pertinent as well. It is common in social protection discourse to see statements about costs that are not well supported by evidence, so capsule summaries of the evidence are provided in chapter 1, in essay 9 on political economy, and chapter 2, in the section titled "Evidence Base for the Costs of Poverty Targeting," on other areas.

In the discourse on the political economy of budgets, taxes, and targeting, "more for the poor is less for the poor" has become something of a mantra, which probably both contains an important idea and simplifies and exaggerates it. This catchy phrase aptly summarizes the intuitive idea that widely shared benefits may garner more political support and thus be allocated higher budgets than more narrowly targeted programs. The median voter theory on which it is based limits voters' preferences to consider only their own net benefit in a single period from different policies. It omits the motivation voters might have about risk management for their own futures or the current or future welfare of extended family, friends, or community members and thus may underestimate the bases of support. The median voter theory even more greatly oversimplifies political decision processes.

Moreover, empirical support for it is rather weak. Consideration of how societies determine their social contracts—the mix of redistribution and risk management and the mix of instruments selected (taxes, spending, or legislative mandates)—is surely important. Yet, consideration of political economy needs to reflect the complexities of the real world and how these may vary from place to place or over time.

A body of evidence from high-caliber impact evaluations shows that work disincentives have been limited in poverty-targeted social assistance in developing countries, smaller than in higher income countries. Many programs have designs that would suggest limited effects—eligibility is not based on income, benefits only loosely correspond to income, updates to eligibility are infrequent, benefits are low, and/or countervailing incentives or services are built into program design. Moreover, the evaluation evidence shows that the transfers can release constraints to work, for example, by paying regularly, the programs may help households buy inputs they need for their farms or microenterprises and thus make work more productive, or the transfers may make it easier for households to afford the resources needed for job search. Although it is more nascent, the behavioral economics literature also suggests various ways in which social assistance may improve work effort or its fruits.

The incremental administrative costs of differentiating eligibility are not zero, but often they are quite low. Data on administrative costs in general and especially those related only to eligibility determination are notoriously scarce and hard to make comparable. However, from the observations assembled here on the costs of large-scale social registries in middle-income countries, which support multiple targeted programs, the costs range between US$1 and US$3 per household in most countries, or in the range of 1–3 percent of the value of benefits channeled through the system. Relative costs are higher in low-income countries with more nascent systems, having not yet amortized start-up costs or reached large scale, about 7–8 percent (see figure O.4). Chapter 4, on delivery systems, suggests many ways in which programs could wisely spend a bit more on administrative costs, in ways that would improve their realization of human rights, lower transaction costs to participants, and reduce stigma.

Transaction costs, stigma, and social cohesion are also concerns. Transaction costs are rarely measured well. Perhaps their most worrisome aspect is in the errors of exclusion that can result when delivery systems are underdeveloped, and those are accounted for already in the coverage/incidence counts. It is clear that better developed delivery systems can reduce transaction costs. The stigma or reduction in social cohesion that may result from focusing benefits on the poorer are the hardest to quantify and weigh in the ledger of pros and cons. Good delivery systems and communications can clearly lower stigma and may be able to influence social cohesion, but these are areas that deserve much more attention in policy and research.

Figure O.4   Ratio of the Unit Cost of a Social Registry versus the Annual Benefit of the Largest Program Served
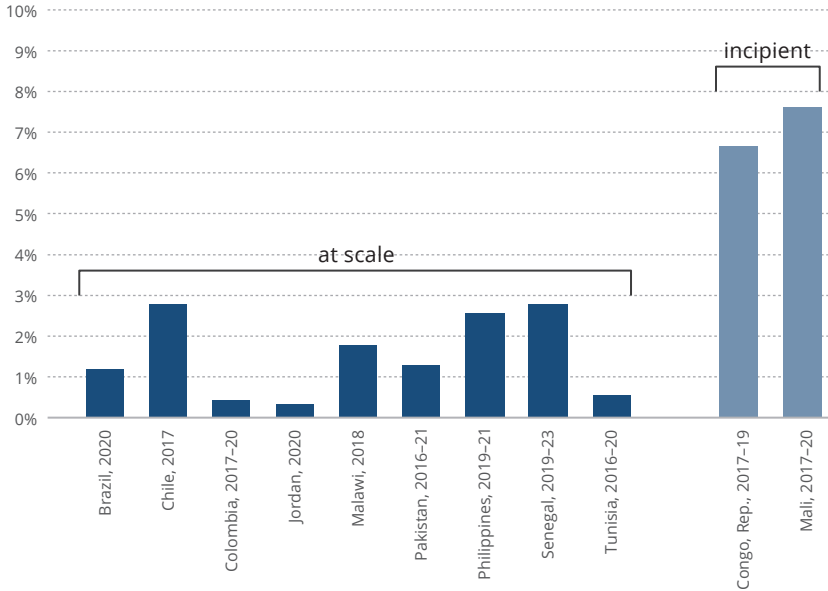
Source: Original compilation for this publication based on Annex 2D.

Message 5. Good delivery systems are critical for delivering all social protection programs, especially targeted programs.

If social assistance is to be a function of government, capacity must be built. Net redistribution requires progressive taxes, progressive transfers, or both. Governments can build the capacity to discern income and assets at (at least) the top end of the distribution, such as when building income, property, wealth, or inheritance taxes to build progressivity on the revenue side. Governments can also build the capacity to discern welfare at the bottom end of the income distribution to administer poverty-targeted transfers funded from sources like excise, value added, or resource extraction taxes/revenues. They can build capacity in administrative registers such as tax systems, social assistance application processes in social welfare offices, or (partial) census sweeps to register information about households in their homes. Ideally, one or both of the registers will cover the too often "missed middle." Governments can locate the associated workforce in central agencies or local governments or harness the power of semiofficial and unpaid community members. Somewhere, capacity must be built, and the better

the capacity is, the better the social assistance can be. Using the costs or dif-
ficulty of building capacity to discern the welfare of the poor as the reason
for preferring universal programs over poverty-targeted ones skips a piece
of the logic. The funding of universal systems also requires building capac-
ity to assess income or consumption, it just builds it in a different place—in
the tax office rather than the social assistance center.

The delivery systems and processes of social assistance programs are
important for the distributional outcomes—for reducing both errors of
exclusion and errors of inclusion. Acknowledgment of the importance of
implementation is not new in the literature, but this book goes deeper into
the topic than many treatments. Chapter 4 reviews the stages of the deliv-
ery chain from outreach through exits from the program, using Lindert et
al.'s (2020) framework (figure O.5), highlighting the main ways in which
targeting errors can occur and how they can be minimized. It discusses how
the delivery chain can be strengthened to allow programs to handle shocks
and the institutional and data systems to support the delivery chain.

Figure O.5   Social Protection Delivery Chain
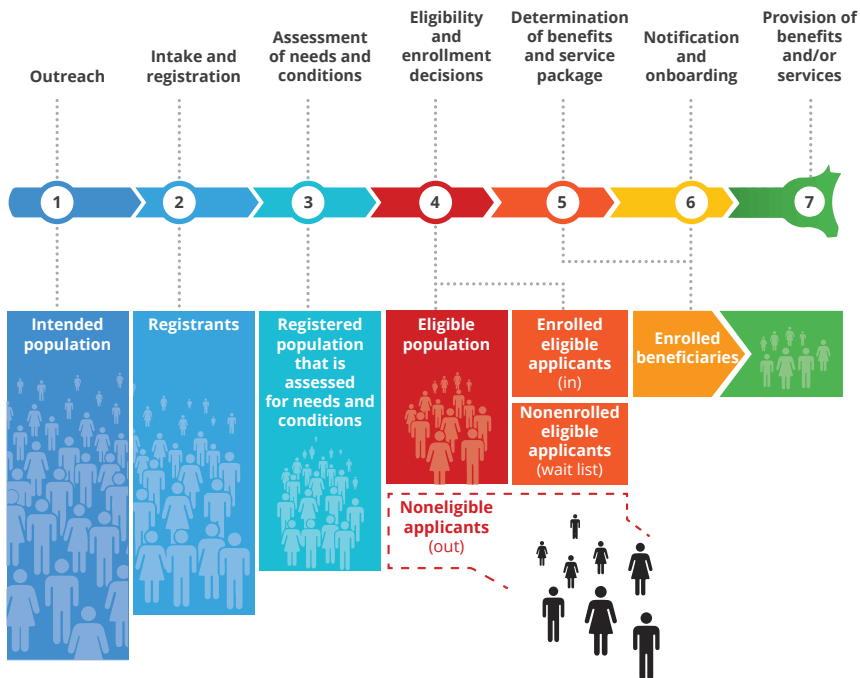
Source: Lindert et al. 2020.

Chapter 4 contains the most focused discussion of the importance of delivery systems for targeting, but the importance of the topic is such that there are echoes in other chapters as well.

The book covers delivery systems even before discussing the choice of targeting methods, to emphasize the importance of implementation of different elements of the delivery chain for improving targeting performance, especially lowering errors of exclusion. No matter how aptly selected the targeting method, and no matter how good the data or inference it relies on, it cannot deliver good outcomes without good implementation of each step of the delivery chain. Indeed, understanding how crucial delivery systems are comes in part from the literature on age-based social pensions and child allowances, which, despite their universal design, struggle with some of the same practical issues as poverty-targeted programs to get to the desired level of inclusion.

Good delivery systems are important for compliance with several of the principles of the human right to social security. Chapter 4 identifies many aspects of implementation that support accessibility, dignity and autonomy, nondiscrimination and equality, inclusion of vulnerable groups, gender sensitivity, and transparency and accountability as they are understood in the right to social security. For example, providing physical accessibility is important for people living with disability, and materials and staffing for various languages as needed are important for nondiscrimination, dignity, and inclusion of vulnerable groups such as indigenous groups, ethnic minorities, and immigrants. Providing clear information on processes can help people know whether and how to apply or appeal, which will lead to high inclusion and be in keeping with the transparency and accountability standards of human rights. Ensuring that all processes are effectively accessible to women is in keeping with gender sensitive social protection. Indeed, a great deal of the bad reputation of targeting with respect to human rights is earned through insufficient delivery systems rather than inherent in the process of eligibility determination. Human rights perspectives can be especially useful in spurring or guiding improvements in delivery systems.

There is room for substantial improvements in the current practice of delivery systems. The following are among the most important:

- To improve outreach and communication so that people who are meant to be served by programs are aware of them and know how to access them
- To ensure low transaction costs (the time, travel, and mental bandwidth of those in pursuit of benefits, in calendar time in queue) and improve the client experience of inclusion and dignity
- To develop dynamic intake processes so that all who are eligible can apply promptly rather than waiting years for the chance

- To develop routine or ongoing recertification or exit processes with a periodicity to match the program objectives and expected dynamics of changes in households' welfare
- To prepare in advance for expectable disasters and crises, with triggers and emergency rules of operation laid out
- To build the client interface systems and capacities to run the programs well, with good governance and convenience for clients
- To upgrade practices in data management and data protections apace with the greater use of technology in delivery systems

Message 6. A range of targeting methods exist; program objectives and the country, social, and political context are likely to influence the choice; and there is no absolute ranking of methods.

The menu of targeting methods is well-established (see table O.1), as are their general advantages and disadvantages. Although the suite of methods has the same names as written about two decades ago, the practice and potential of each is changing as new data, new technology, and new capacities and expectations push them to evolve. Issues on the choice of method are taken up in chapter 5.

Chapter 5 briefly reviews the patterns of which targeting methods have been used where and why. In general, means tests and hybrid means tests are predominant in high-income/high-formality settings and Europe, although they were used in some upper-middle-income countries, such as Brazil, China, and South Africa, even before data systems were developed for verification of means. Proxy means testing developed in the relatively high-inequality and high-capacity countries in Latin America but has spread far beyond. Some view proxy means tests as desirable for bringing data-driven, replicable, technocratic processes to replace previously highly politicized alternatives. Others view them as anathema due to their in-built statistical errors and perceived opacity. Still others view them as an imperfect but realistic approach where other options are unlikely to succeed. Lower income countries use a mix (and often a combination) of proxy means testing and community-based targeting. Community involvement in eligibility determination is highly varied. In some places, communities have a large role in the actual decision making; in others it is less so, although communities may play important roles in outreach or data collection. Geographic targeting is used in various ways—sometimes to select the areas in which a program will work, sometimes to ration the caseload across the areas served, and sometimes both. It can also be important to focus on where administrative resources should be dedicated to improving

Table 0.1   Common Targeting Methods

All programs have an element of self-selection:

Self-targeting. *Implicit.* All programs are implicitly self-targeted in that individuals decide to participate or apply if they consider the package of benefits and program rules acceptable. As discussed in chapter 2, many elements of program implementation that affect transaction costs and stigma affect that calculation and degree of self-targeting. *Explicit.* Some programs also have explicit design features to promote differential take-up across who the population. Public works programs pay low wages for short periods, with the idea that those with better employment will not participate. Food subsidies or rations may feature nutritious staple foods that are a larger share of the diets of the poor than the less poor, and sometimes the less prestigious versions—broken rice, coarse flours, unattractive packaging.

Some programs operate by defining broad categories of eligible households:

Geographic targeting. In the strict version, a program selects/serves only those in a defined geographic area; in a more moderate version, a program allocates rationed caseloads to different areas based on spatial variation in need. Either version requires the ability to delineate boundaries, which may be clear for units of political representation (state, district) but less so for smaller areas (village, neighborhood). And it may entail some requirement of "belonging" or duration or formal registration of residency.

Demographic/categorical targeting. This applies when benefits are granted to people according to their membership in fairly easy-to-observe categories. The categories that are most commonly used and easiest to observe are based on age, civil status, and gender, though programming may also be directed to veterans on a categorical basis. Ethnicity is occasionally used, as in affirmative action programs.

Some programs seek to distinguish the welfare of specific households:

Means testing (MT). When household's income and/or assets determine eligibility, often these are verified against independent sources.

Hybrid means testing (HMT). When a significant part of the information on the family or household's socioeconomic condition can be verified against independent sources, and the other part needs to be imputed or predicted.

Proxy means testing (PMT). When information on the family or household's socioeconomic condition needs to be estimated/predicted based on (mostly) observable sociodemographic characteristics and economic assets because verification of socioeconomic status cannot be performed.

Community-based targeting (CBT). When community leaders or members use information known to them from day-to-day living in the community to guide or choose who should be in or out of the program. As part of this assessment, the community may be guided to use wealth ranking or household economic analysis (HEA) techniques or similar techniques.

Some people use the term **affluence testing** to refer to programs with the eligibility threshold set quite high up the welfare distribution rather than toward the bottom end. Affluence testing is not a method per se in that it isn't about what information is used to determine needs. It is often associated with means testing but can apply to HMT and PMT as well.

Some programs seek to ration without further ranking or comparison of need:

Public lottery. When a random process is used to ration spaces among eligible applicants in an oversubscribed program. In a sense, this is less of a targeting method as such and more an additional way to ration selection.

Source: Original compilation for this publication.

delivery systems. Demographic targeting is another mainstay with many programs for children, the elderly, or their families.

The literature is not definitive in ranking among targeting methods, as context and capacities shape the possibilities, nor is it definitive in matching methods to contexts, as preferences and history shape choices. Although there are some patterns as described, there is enormous variation in implementation and virtually every combination of methods has been used somewhere, sometimes in seemingly unlikely places.

The question of whether to use simpler methods, such as self-targeting, geographic targeting, or demographic targeting, or to develop household-specific methods must be based on "fit for purpose" as well as context and capacities (figure O.6). Using geographic targeting to select only some areas in which to work may fit well with geographically delineated natural disasters, but it occasions large errors of exclusion for poverty-oriented

Figure O.6 Factors to Consider in Choosing a Targeting Method
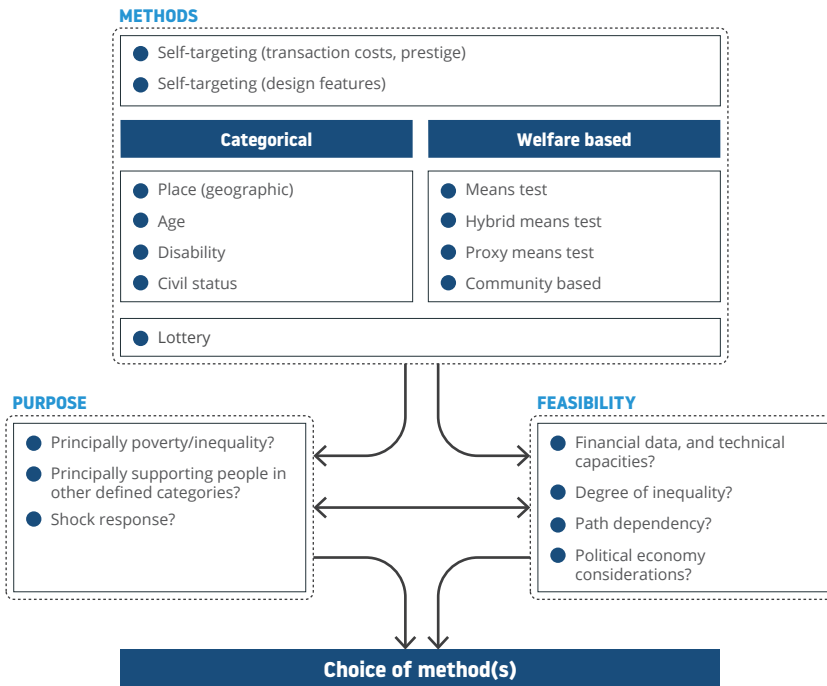
Source: Original compilation for this publication.

programs in normal times. Demographic targeting for some purposes is axiomatically ideal; but for poverty-related purposes, it will be inexact although possibly pragmatic. Further, the fit for purpose can vary by program within a given country. For a school lunch program, geographic targeting to poor areas, possibly excluding any categories of schools (private or upper secondary) that serve students who are less poor, may be appropriate, whereas for a last resort income support program, some household-specific targeting may be used.

For household-specific targeting, there is a fairly clear order of preference, although sometimes the context narrows the choice set considerably. In some countries, means testing is feasible and with no inbuilt statistical errors, it is easily adopted as the first choice. This range of countries may be extended by hybrid means testing, which may have some errors in the imputation of informal income, but the imputations affect only some households and some of their incomes are still lower than in many methods. The range of countries where such methods are applicable is increasing with the secular trend in data availability and may be applicable in still more countries for programs where the eligibility threshold is set high. In some countries with high informality, means testing or hybrid means testing will not be able to distinguish the poorest, but it may be able to rule out the wealthiest. However, in many developing countries, the degree of informality implies that means testing, or even hybrid means testing, will not be very accurate and so those choices are often deemed to be off the table. In these cases, proxy means testing, community-based targeting, or some combination becomes the common option. In many places, the community has long been a part of targeting processes and, although its role may change from full-out decision making to more supportive roles in outreach, data collection, and monitoring, the community will maintain a degree of involvement due to path dependency. Conversely, in some settings, the degree of community cohesion may not allow community-based targeting. This may be true in urban settings where density and mobility (in both residence and where time is spent during each day) are so high that people do not know their neighbors well, or where geographical communities are socially divided by ethnicity or conflict. Still, proxy means testing, community-based targeting, and combinations are methods that are still on the table and used in a large share of developing countries. Where these are insufficient or undesired, some rationing, such as by geography, demography, other observable characteristics, self-targeting, or even lottery, may still be an option.

In countries that have a well-developed method of household-specific assessment, multiple programs that use household-specific assessments may use the same means of assessment, although with possibly varying

thresholds or ancillary criteria. Using a common process or shared social registry as the entry point for multiple social programs has benefits and risks. By providing a shared format, it harmonizes the information collected and can add coherence across social policy. By serving as a common portal, it can lower the costs of application as a household may have to apply only once to receive multiple sets of benefits, or at least receive cross-referrals that improve their knowledge of programs from which they may benefit. Similarly, shared registries may lower the total administrative effort governments put into outreach, intake, and registration, and by uniting the efforts across various programs, they may be able to amass resources and gravitas to do the work well. However, in concentrating provision, a common process also concentrates risks as any failure in outreach or process affects not just a single program but many. The heavier is the use of the registry in social policy, the more important it is that it be dynamic, inclusive, and accurate.

The use of multiple targeting methods is common, but it is not necessary and sometimes not ideal. For example, although categorical methods can help prioritize when budgets are limited or sequencing of the rollout of a new program is required, they are guaranteed to exclude poor people as poverty affects some people in all places, all ages, all genders, or other group definitions. So, if a country has developed the capacity for a household-specific system, does a categorical system add value?

Message 7. There are better and worse ways to implement each targeting method, and lessons have been learned over time.

No matter what targeting method is chosen, its application must be carefully customized. There is no guarantee that what worked well in one place will work the same in another. Although general principles may carry through from one context to another, customization will be needed to account for country- and program-specific details such as the purpose and design of the program; availability of data; capacity of the delivery system; characteristics of the population of concern; and weight put on the different desirable but sometimes conflicting traits of low errors of exclusion, low errors of inclusion, and low costs of all the various sorts. Customization includes the definition of the assistance unit, the thresholds used for each program, the roles assigned to each institutional actor, and the like. Customization involves the detailed decisions involved in moving from abstract concepts to implementation, as described in chapter 3; the delivery system, described in chapter 4; and each method's know-how, described in chapter 6.

While mindful that customization is needed, there are a few rules of thumb to guide planning and assessments of practice.

For some facets of eligibility determination, more is probably better than less, although of course tempered by cost. The following are some examples:

- More care to outreach, communication, mechanisms for grievance redress, human-centered design, and client dignity is better than less.
- More care in building capacity at the local level yields benefits, including for community actors who may be involved in community-based targeting, outreach, or monitoring.
- Open registration is more desirable for most kinds of programming, compared with only periodic, especially infrequent, registration.
- Greater coverage of foundational identification (ID) will facilitate many processes. Social protection programs can help facilitate access to foundational IDs but will need to have work-arounds for those still lacking them.
- More data are usually better than fewer, more recent data better than older data.
- A greater degree of interoperability among various government registers that help to define the assistance units (ID agencies or civil registries) and their welfare (records of income or social security contributions, and land, automobiles, payments for government services, or receipt of government benefits) is helpful as long as due data privacy and security provisions are in place and respected.
- More regular and multimodal monitoring of implementation and results can allow faster adjustments and improvements.

For some facets of eligibility determination, there may be a sweet spot between too little and too much. The following are some examples:

- Programs that are smaller than the target population will have exclusion errors by design. A program needs to be at least as large as the population for which it is intended and preferably a little larger; being somewhat larger than needed will likely reduce exclusion errors while those incorrectly included are unlikely to be very wealthy. At the same time, being much larger than needed is costly and begins to include those who are not part of the population meant to be served.
- For recertification, very high frequency may raise costs and errors of exclusion unduly, but excessively long periods without reexamining eligibility will surely result in errors of inclusion. If budget/places in the program are rationed, lack of recertification will lead to errors of exclusion as well.
- Means testing, even hybrid means testing, requires building a reasonably comprehensive data system to measure income and assets, but demanding too much detail can push clients into fraud, disincentives for work, or withdrawal from the program.

- For proxy means testing formulae, many countries use simple, single-model ordinary least squares, but better results might be obtained from a little more complexity. In modeling, the use of quantile regressions and auxiliary data (big data) at the geographic level is likely to improve prediction. Whether more complex methods of machine learning will pay off is less clear or generalizable. Likewise, having a simple national model may be less accurate than having a suite of models for different areas (metropolitan areas, towns, and rural areas) or administrative units (states or provinces), but having many different models may require more data than are available, or it may introduce practical issues for implementation and communication.
- Some systems are designed in a way that requires greater capacity than the program or country can muster and might be better simplified; other countries fail to make improvements that are seemingly within their reach.

Message 8. Income dynamics and shocks are significant and pose difficult challenges for eligibility determination processes; some targeting methods are more agile than others.

Even in normal times, the dynamics of welfare and poverty are considerable; shocks can dramatically amplify this. Chapter 3 provides examples of how fluid poverty is in many countries and regions, often with the number of transient poor being at least the same as the number of chronically poor. Volatility in income can be driven by good or bad harvests and seasonality of work in services or by job loss, illness, or an accident. Natural disasters, climate change, economic crises, and pandemics can disrupt livelihoods, at least temporarily, sometimes permanently and for many people at once.

When trying to forestall the long-term negative consequences of shocks, whether idiosyncratic or covariate, the speed of assistance can be of utmost importance. The logic is intuitive and substantiated in the formal academic literature. If assistance is to prevent a negative coping tactic, it must be timely, before a family's baby becomes malnourished, before it withdraws a child from school, marries off a child bride, sells its assets, racks up high-interest debt, or loses its home, workshop, or land. Each such coping tactic can be very difficult to reverse, ratcheting down the individual or family's welfare for years or the rest of their lives. Assistance (usually temporary) can help prevent such losses.

The recurrence of shocks and crises and the premium on swift response pose the challenge of how social protection systems can be adequately flexible and dynamic. Given the focus of this book on eligibility determination, it considers this element among the wider aspects of adaptive social protection (building resilience, ensuring adequate financing for crisis response, and building institutional frameworks and capacity). The conceptual and

measurement issues are treated in chapter 3. Several of these topics pertain to improvements or adaptations to delivery systems and so are treated in chapter 4. The pros and cons of the different targeting methods for emergency response are treated in chapter 5, and the how-tos are presented in chapter 6, including the use of new data and technology.

Shock responses require thinking through who gets the priority for assistance—those who were poor even before the shock? Those made poor because of it? Those with large losses even if they remain above the poverty line? In the ideal, all three groups would benefit from social policy but likely via different sorts of responses and for different reasons.

- Helping the chronic poor after a crisis may not be sufficient, but it is this group that may most quickly have to resort to negative coping tactics and so they should get first priority. This is often relatively feasible since it is by far the simplest and fastest social protection response to issue an emergency top-up payment to people who are already in some social assistance program. Often, expanding ongoing but low-coverage programs is the next fastest option as there is a base of systems and personnel from which to start.
- At the same time, a crisis response beyond helping the already poor may be needed to reach the new poor or those who have suffered significant losses. Often relatively broad, flat (or minorly customized) benefit designs are used for crisis response programs. This simplifies eligibility decisions and balances poverty reduction and risk management goals. The government may also mandate or facilitate insurance programs to help cover risks ex ante.

Some targeting methods lend themselves more easily to handling some sorts of income dynamics or shocks than others:

- Geographic targeting fits well for natural disasters, which are usually spatially delimited, but it is not very apt for economic crises, which usually affect all areas of a nation.
- Demographic/categorical targeting is not a natural match for covariate shock response per se—no one's age changes in response to a shock; natural disasters do not strike only those of some ages; and economic disasters hit workers/those of working age more directly and their dependents only indirectly. Nonetheless, top-up benefits to beneficiaries of demographically targeted programs may be a way to get money out quickly, especially where coverage of such programs is high. Of course, children are so biologically vulnerable that it is always important to protect them. Demographic targeting is something of a recognition of the idiosyncratic shocks that come as families move through the life cycle. Child grants and social pensions help cushion changes in the dependency ratio within families.

- Among household-specific methods, means testing and hybrid means testing, which rely to a large extent on data from interoperable government systems that maintain high-frequency data, can be fairly agile in responding to idiosyncratic and covariate shocks. This can be especially true for eligibility determination that draws on monthly or biweekly data on contributions to social security systems or income tax withholdings as these reflect changes in wages or formal employment in short order. Eligibility that is based on longer term measures, such as annual income tax information or holding of assets, is less responsive.
- Proxy means tests are basically calibrated to reflect long-term welfare, traditionally have been based on characteristics of families and their assets that change slowly, and have tended to be used with data updated only every few years, so these tests are less able to be shock responsive. Some recent innovations merge measures of exposure to weather or geophysical shocks with more traditional proxy means testing data to attenuate the problem, while faster-changing data such as phone data may offer some promise.
- Community-based targeting assessments seem to be able to pick up how households are affected by various shocks, but they require updating after the shocks hit and thus may take some time.

Many actions that are important for preparing social protection systems to be responsive to shocks are also important for moving to USP in general and vice versa. Improved coverage of the chronically poor in normal times is important for USP; it also builds resilience before shocks and makes top-up programs feasible. Such full coverage requires a continuously open enrollment process, adequate base financing, and enough flexibility to ensure that entitlement obligations are met, at least through normal swings in need. It thus provides a base of response in times of crisis. High coverage of foundational IDs (especially electronic identification [eID]) can help provide links to many sorts of data and facilitate some rapid (possibly simplified) eligibility assessments. Foundational IDs coupled with extensive financial inclusion also facilitate quick payments. Building out the insurance part of social protection systems serves both USP and resilience.

Message 9. Advances in technology—Information and communications technology, big data, and machine learning—offer the promise of significant improvements in targeting accuracy but are not a panacea; better data may matter more than greater sophistication in inference.

A key element of targeting is using data or inference to discern different degrees of poverty or vulnerability. Changes in technology and the availability of new data always excite hope that these will make targeting more

accurate or easier. Deeper discussion of these issues is concentrated in chapters 6 and 8.

Improvements in the availability and use of traditional government-held data have been and will continue to be a driver of improvements in the ability to observe welfare and target, especially potentiated by the increasing use of foundational IDs and eIDs. Increasing the coverage and quality of such data systems and the ability to conduct data matching are helpful for most methods of targeting, especially in facilitating means testing and hybrid means testing (and a move away from welfare proxies). Improvements in the scope and quality of traditional government data—on taxes (payroll for firms, sales for value added tax, personal income, and property such as land or automobiles); on fees for government-provided services (especially utilities and border crossings); and on the use of government-provided services targeted in various ways (other social assistance programs and sector-specific preferences or privileges such as fertilizer discounts and fee waivers for any government-provided services)—help make welfare observable. Governments have held such data for many years, but their use in eligibility determination may be improved with attention to the technical details of definitions and data structures, which make it easier to match among data sets, and with attention to policy issues of data privacy and data security that regulate the legality of doing so. Many countries are rolling out or extending coverage of foundational ID systems and often upgrading to eIDs, which will make much more data matching feasible in proximate years as the eIDs become the keys for matching on more government-held data sets. Drawing on the integration and interoperability of such data system matching in eligibility processes can reduce the need to collect data again and again and can facilitate cross-referral processes from one program to another, which can lower transaction costs for applicants and governments alike.

Where welfare is difficult to observe directly, targeting methods try to infer it from observable proxies; whether the proxies are new or old, they need to be highly correlated with welfare. Nonadministrative big data— such as from satellite imagery, mobile phones, and social media—and machine learning are expanding the data and techniques for this at a dizzying pace, although they remain largely proxies for welfare rather than direct measurements per se.

Although they are often still proxies, big data have the advantage of not requiring household-specific data collection by the social assistance agency via lengthy intake interviews or (partial) census sweeps as they are generated by other government or private processes. However, the social assistance agency must acquire and use them. Thus, they offer the prospect of being cheaper and faster for eligibility assessment, allowing not only rapid program start-up, but also more frequent reassessments as conditions change.

Big data are already being combined with traditional data to improve poverty maps and help predict which households and areas are more at risk of natural disasters. Administrative data have long been used alongside traditional data to improve poverty maps; newer big data can similarly be incorporated. Moreover, historical data on localized natural disasters and drought combined with realized household poverty outcomes can be used to predict which households will be at risk in the future. Such models can be used to target the poor or vulnerable for covariate risk-mitigating social protection programs or public insurance schemes, helping administrators to manage covariate shocks.

In a crisis or data-scarce environment such as postconflict, using big data for determining eligibility may be one of the only options and an appropriate one. Big data can fill a gap when traditional data are not available, as in many poor or fragile countries or in postconflict settings, or when the data are not current, as in a crisis. In such circumstances, the ability to target much-needed assistance is vital.

Whether big data will replace the need for traditional data for eligibility assessment depends on whether the challenges arising from their newness can be fully understood and solutions crafted. Some of these challenges are well-known and require as complement more traditional data, such as for training and assessment. Some of the challenges are well-known and require care in implementation, such as avoiding bias in models. Other challenges require new thinking and new research, such as matching the unit of assistance and understanding impacts on behavior.

- *Ground truth for training*. Big data are increasingly used to generate poverty maps at a fine-grained level and where traditional data do not facilitate them. Their viability still relies on accurate ground truth training data, that is, household surveys with direct income or consumption measures. In their absence, many big data–based maps use survey data such as the Demographic and Health Survey series where household welfare is not directly measured but instead estimated with proxies; in essence, big data proxies are often used to model another proxy rather than the direct measure of interest. This is a limitation for traditional poverty maps using census data as well, but big data do not overcome this.
- *Ground truth for assessing*. Proper assessments of different big data maps—from satellites, call detail records, or social media—compared with traditional maps and survey data with directly measured household welfare are still needed to understand whether the big data maps are more or less accurate than traditional methods, and thus whether they should be preferred to traditional maps or only used when the latter are unavailable.

- *Avoidance of bias in prediction*. Machine learning models use big data to learn and predict. When the data they train on do not represent the whole population, the model predictions can be biased. For example, early face and voice recognition models are much better at predicting white males than nonwhites or females. Careful checks need to be put in place to ensure that eligibility assessments do not disadvantage particular subpopulations; the marginalized groups of interest to social assistance policy may often be exactly the ones missing from big data.
- *Unit of observation.* Eligibility is often at the household level, while big data rarely are. Even newly fine bore geospatial analysis remains at a grid level rather than being household specific. Data from call records may pertain to the subscriber identity module (SIM) card or phone number, of which an individual or household may have none, one, or several, separately or shared across individuals or households. These issues add a level of complexity to the use of such proxies.
- *Data access and use.* Many big data are privately held. What regulation or incentives it will take to make such data available to core government functions on an ongoing basis (beyond just in a crisis), or what is socially acceptable (for government to access and for what purpose), is mostly still to be worked out.
- *Incentives.* Just as there has been great concern that using more traditional administrative big data might generate undesirable labor incentives, it will be of interest to learn whether the use of phone data or social media for eligibility determination will change behaviors in ways that reduce people's welfare or reduce the accuracy of the proxy.

More sophisticated inference—machine learning—is probably less important for better targeting outcomes than more and better proxies—big data. The small literature exploring the use of machine learning algorithms finds that the algorithm that produces the best proxy means test depends on which metric is being used to evaluate and how the scoring would be implemented in practice. It also generally finds that the improvements in performance are relatively small compared with traditional models. Thus, it is not clear that the complex analysis required to determine which is the best machine learning model in a particular country context for a particular program objective and design is worth the improvement over more traditional models. Moreover, the increase in opacity—a black box on top of a black box—may concern policy makers in some countries, although machine learning–based proxy means testing models were recently adopted in Colombia and Costa Rica; new visualization tools can also help make the models more intuitive. Where significant improvements in machine learning models have been identified, the improvements are driven more by bringing more proxies into the model—whether administrative data or

"feature engineering" (developing new variables from within the traditional data itself)—than the choice of model itself.

In the end, the use of new forms of big data and sophisticated inference should be understood as an interim step in the transition to measuring welfare and eligibility directly. Most nontraditional big data remain proxies for the underlying welfare that it would be preferable to measure directly to determine program or benefit eligibility. It remains that analysts resort to machine learning or traditional regressions to estimate the underlying welfare from the proxies, traditional or new. New data and new techniques may help reduce the inherent modeling error of proxy means testing, but such errors will remain. It is expected that as more and bigger data become available on which to train machine learning, the combination will soon become increasingly common. Yet, ultimately, an improved proxy means test is still not a substitute for direct measurement of most or all income nor for the need for interoperability and data integration.

Increasing use of new data and inference will not replace the need for humans in all parts of the provision of social services. New data and inference may improve the accuracy, increase the speed, and lower the cost of eligibility assessments. They may help lower transaction and administrative costs for some clients and some functions. However, some clients may need human social assistance workers to help overcome issues of information, agency, or the digital divide. Some processes, such as grievance redress or referral from income support to social services, may benefit from rapport built between the social assistance staff and clients.

Message 10. How countries target is often and should always be a dynamic story.

In many countries, efforts to target social assistance have evolved over time—often improving aspects of delivery systems or data collection on a continuous or recurrent basis, sometimes improving formulae and data use, and occasionally evolving from one targeting method to another altogether. Stories should be dynamic where new programs, problems, or heightened expectations demand attention and as new capacities, new data sources, and new computing power move the frontiers of what is possible. Sometimes there are reports of government administrations of different political orientation focusing on different sides of the targeting problem—with one putting more emphasis on reducing errors of inclusion and another on reducing errors of exclusion. Where taken in alternating turns or by different levels of government, with balance and technical quality, both emphases can contribute over the years to improved programs and impacts. There have also been occasions of stagnation when countries or programs have stalled in their efforts at improvement. These are a reminder

that secular changes in technology do not ensure progress; progress is always the result of political will and administrative effort, resources as important as the budget in producing good social policy outcomes.

Notes

1. The Global Partnership for Universal Social Protection brings together the World Bank and the International Labour Organization with the African Union, Food and Agriculture Organization, European Commission, HelpAge, Inter-American Development Bank, Organisation for Economic Co-operation and Development, Save the Children, United Nations Development Programme–International Policy Centre for Inclusive Growth, United Nations Children's Fund, and others, along with Belgian, Finnish, French, and German cooperation (https://www.usp2030.org/gimi/USP2030.action).
2. Less any transactions for having applied to or psychic costs of having been excluded from the program.
3. The estimates are simple in the sense that they do not consider behavioral responses to the programs.
4. http://www.worldbank.org/aspire/.
5. ASPIRE focuses on data from traditional International Bank for Reconstruction and Development and International Development Association client countries, so its coverage of high-income countries is not the full set for the world (for example, the earliest industrializers, Australia, Europe, Japan, North America, and so forth). Thus, the higher income sometime-borrowers from the World Bank for which ASPIRE reports survey data for the period since 2014 are Chile, Croatia, Panama, and Uruguay.
6. Some of these programs may intend to cover more than just the poorest quintile; the matching of program size to that of the target population is a point taken up in Message 7. This message instead makes the point that programs that are very small will never have a significant impact on covering the poor regardless of how well their eligibility screening works, while those that are very large will cover the poor but the poor's share of the total benefits will be limited.
7. Two conditional cash transfer programs are observed in Panama's 2018 Encuesta de Mercado Laboral: Red de Oportunidades, which covers 5.7 percent of the total population, and SENAPAN, a smaller program covering less than 1 percent of the population. Panama also has two additional programs targeted to poor individuals: Programa Ángel Guardián (for people with disabilities) and B/.120 a los 65 (for adults 65 years and older who do not receive a contributory pension). Although Panama's government considers the programs as conditional cash transfers due to the inclusion of conditions, ASPIRE classifies them as social pensions since their conditions are not related to investments in human capital, such as school attendance, immunizations, health checkups, and so forth. See annex 2C in chapter 2.

References

Lindert, Kathy, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Nishikawa Chávez. 2020. *Sourcebook on the Foundations of Social Protection Delivery Systems*. Washington, DC: World Bank. http://hdl.handle.net/10986/34044.

World Bank. 2012. *Resilience, Equity, and Opportunity*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/443791468157506768 /Resilience-equity-and-opportunity-the-World-Banks-social-protection-and -labor-strategy-2012-2022.

World Bank. 2021. "Social Protection and Jobs Global Compass." PowerPoint, World Bank, Washington, DC.

1

Targeting within Universal Social Protection

Margaret Grosh

Within the diversity of social protection systems, social contracts, and the mix of institutions and policies used to achieve them across countries, three constants shape the discourse and practice around prioritizing those in need.

First, there is a strong consensus around the determination to reduce poverty and inequality and a drive toward universal social protection. That consensus is reflected in many national policy statements and even some constitutions mandating universal social protection. The goal of universal social protection has been codified as part of the Sustainable Development Goals to be met by 2030 and supported by a long list of international organizations, including the World Bank.

Second, it is a fact of life that hundreds of social programs around the world differentiate eligibility and/or benefits in various ways. Nearly every country has at least one poverty-targeted social assistance program. Many countries have multiple programs in different parts of social policy that base eligibility or differentiate benefits according to welfare levels,[1] and often one or more of these are high-profile flagship programs. Many countries have special programs to support children and the elderly, because they are deemed social priorities, more likely to be poor, or both. The unemployed may benefit from unemployment insurance, and those struck by natural disaster may benefit from assistance initially to sustain them and eventually

to help them rebuild housing or livelihoods. Productive inclusion programs seek to raise the level or decrease the variability of the incomes of the poor. Active labor market policies are usually focused on those with greater barriers to (re)employment. Thinking more broadly—about those lagging in education or without access to health care or essential utility services— many more programs exist to direct various social policy efforts to members of these groups and improve life outcomes for them. These "targeted" programs assist in achieving the goals of universal coverage and sit next to universal programs in broader social policy, with the mix of universal and targeted programs varying from country to country.

Third, the job of targeting or prioritizing among individuals or groups is fraught with conceptual and practical difficulties, has errors and costs, and has many criteria and metrics by which success or lack thereof can be gauged. Thus, the issue of whether current practice is acceptable, can be improved upon, or should be abandoned recurs in instance after instance.

The tension between and within these three constants makes the choices around whether and how to target those in need of different facets of social protection a perennial topic in social protection policy discussions. Responses have varied from place to place and over time, depending on each country's resources, the challenges it faces, and the specific progress and gaps in its social protection system and wider social policy. The varied responses have generated a rich set of global experience from which countries can draw as they review and renew their progress toward universal social protection.

The economic trauma accompanying the COVID-19 (Coronavirus) pandemic has brought higher visibility to social protection. The crisis is nearly unprecedented in its ubiquity, but the issues it highlights are not new: social protection coverage has always been partial. There are gaps in social assistance of various intensities for the chronic poor and the informal sector uncovered by labor protections; delivery systems are often too rigid to respond fully to shocks whether from natural disasters or economic turbulence; how to use social protection to connect poor or unemployed people to better independent incomes is an enduring issue; and financing is a constraint to realizing the full vision of universal social protection. The dramatic scale of response to the current crisis highlights both what can be done when exceptional effort is made and the extent of the need for improvement.

This chapter is the first of three chapters on the tensions between the idea of universal social protection and the idea of targeting specific programs to specific groups. The chapter covers the tensions between universal social protection and targeting in an abstract sense. Chapter 2 takes up the empirics of the errors and costs of targeting those more in need of social protection. Chapter 3 moves from the abstract to more concrete issues in

defining priority groups. The discussion of metrics is embedded throughout the book, with the most in-depth discussion in chapter 7. This chapter is constructed as a series of essays. The first two take up the issue of how selectivity fits conceptually and practically within universal social protection; the next essays take up the principles of targeting and extensions around them; then issues of tax, political economy, and human rights are treated.

Essay 1: Where Does Targeting Fit Conceptually within Universal Social Protection?

International Labour Organization recommendation 202,[2] one of the iconic statements on universal social protection adopted by 185 countries in 2012, acknowledges the many different objectives of social protection, inter alia, as an important tool for preventing and reducing poverty, inequality, and social exclusion; promoting equal opportunity and gender and racial equality; managing risks; and realizing social and economic rights. These ideas carry through in the social protection strategies and definitions of many international agencies that are influential in the field, including the World Bank (see Jorgensen and Siegel [2019] for a short comparison).

The World Bank's Social Protection and Labor Strategy is shaped around resilience, opportunity, and equity. The strategy paper says the following:

> Social Protection and Labor programs directly improve resilience by helping people insure against drops in well-being from different types of shocks and equity by reducing poverty and destitution and promoting equality of opportunity by building human capital, assets and access to jobs and by freeing families to make productive investments because of their greater sense of security. At a macroeconomic level, well-functioning social protection programs are central to growth-promoting reforms. (World Bank 2012, i)

Differentiation of benefits is engrained in the contributory social insurance that is the classic social protection pillar associated with resilience. The benefits of old-age and survivors' pensions, disability insurance, and unemployment insurance commonly have an element of "replacement wage" or "share of earnings" in their formulae.[3] The differentiation of benefits is part of the guarantee, and it is relatively easy to accomplish since wages are observed as part of the payroll tax that finances the insurance. Health insurance differentiates benefits on the basis of the severity of illness. And so on with insurance in allied fields. In crop or weather insurance, which can provide protection to incomes, or property insurance that protects assets, the payout depends on the degree of measured or approximated loss.

Differentiation of eligibility and/or benefits happens in some but not all the aspects of the labor programming and regulation that are part of the opportunity pillar. Efforts to improve the employment prospects of youth or the unemployed may conduct profiling to assess employment prospects and then give greater attention to those assessed as having the most barriers to overcome, for example, in terms of education, training, experience, work habits, health, or the logistics and costs of dependent care or transportation. Legislated labor protections are usually the same for workers of equivalent contractual status, perhaps differentiated by tenure, but different contractual statuses ("contract" or "short-term" employees, part-time workers, and so forth) may have different protections.

Differentiation of at least net and often gross benefits is integral to social assistance and the equity pillar. To reduce poverty or inequality, the net benefit of transfers and the taxes that support them must be positive for at least the poor, and sometimes the thresholds for net transfers rise markedly up the income distribution. This can be achieved through various designs—from flat and universal benefits supported by imposing taxes more on the nonpoor than the poor, or from transfers that are differentiated by welfare in eligibility and/or benefit level, highest for the poorest.

Many social protection programs blend elements of assistance and insurance, and these often involve targeting. Families suffer many risks for which they do not hold insurance or make ex ante payments (through a variety of failures of insurance markets, myopia, or poverty). Public action may help cushion the losses ex post with funding from noncontributory sources, usually general revenues. The payments are often calibrated to losses or poverty prevention or a mix of the two.

Blending of the assistance and insurance functions may be an increasing trend. Indeed, it is a central recommendation of World Bank thinking about how to improve risk management in a world with significant informality (Packard et al. 2019; World Bank 2019). Blending can be done in various ways: providing subsidies for insurance premia for the poor or vulnerable, providing social protection irrespective of the form of employment, increasing the coverage of social assistance, or providing general revenue–based rather than payroll tax–based financing for benefits. Blending assistance and insurance poses some questions on how to measure coverage on the way to universal social protection since the two subsectors have traditionally been measured in different ways (see box 1.1).

BOX 1.1

How Are Coverage and Universal Social Protection Measured?

In some fields of policy, the objective is equality of outcomes, and that requires equality of inputs. In elections, the goal is for everyone to have one vote and only one vote. In health, the goal is for everyone to have the number of vaccines it takes to produce immunity to the given disease, which varies by disease and formulation of vaccine but not usually by individual. The goal of democracy or immunity takes an action, and the same "dose" for everyone yields the same outcome.

In many spheres, equality of outcomes or at least everyone reaching a minimum standard is desired, but it takes different inputs to achieve that. For example, the Sustainable Development Goals contain the goal of every child learning to read by grade 2. That takes physical access to schools for everyone, which will be more expensive to provide per capita in remote areas than in urban areas, and it may take scholarships or cash transfers to help poor households with the implicit costs of schooling, glasses for children with vision impairments, special teaching techniques for children with learning challenges, and so forth. The universal policy of free schooling is supplemented with actions focused on smaller groups of children for whom the universal policies are insufficient. Thus, in the successive actions—for multigrade or tele-schooling or boarding schools for children in the most remote areas, scholarships for poor children, glasses for children with vision impairment, and discerning which children have learning impairments and need special instruction—differentiated eligibility or services are used toward the goal of universal education. Even in the voting and vaccination examples, although the goal is one vote for each adult or one inoculation for each child, that is, "the same dose," it may take much more active and costly outreach to connect some people to the voting booth or vaccination site. The poor, those in remote areas, the least educated, and those of socially excluded ethnicities may need special efforts to inform them of the value and safety of voting and vaccines and to get services close enough to them to ensure universal coverage.

Measuring progress toward universal social protection is somewhat more difficult than for the voting, vaccination, and reading examples cited. Universal social protection requires coverage, but

how is that defined? In health insurance, the goal is to provide coverage to all so that *in the event* people fall sick, they can obtain health services. For contributory pensions, unemployment, or disability insurance programs, coverage is used in an analogous way and measured based on inscription or contribution to the financing pool. In most periods, people who are covered by such insurance will benefit from the *guarantee* or promise of help when needed (should they get sick, become disabled, or reach a pensionable age), but not necessarily from a *payout.* For social assistance, in contrast, coverage is often interpreted as *receiving an actual transfer*. This is quite different, and it is a critical issue to clarify given the implications for universal social protection. For instance, if a country has a guaranteed minimum income program that provides cash when incomes fall below a threshold, the social insurance interpretation would be that—as in the case of health insurance or pensions—everyone is covered independently of the event occurring (that is, income falling) or whether they are currently in receipt of a payout. Those who are covered would be the whole population, which is usually severalfold greater than the roster of recipients at any point in time. A guaranteed minimum income is universal in insurance terms, but it is poverty-targeted from a social assistance perspective.

And, of course, coverage is not the only requirement for universal social protection. Coverage must include benefits that are adequate and fit for the intended purpose. In health, the degree of resources needed for each person in each time period will vary according to whether they are ill or the severity of their illness. Similarly, in social protection, people will need different degrees of support depending on their exposure to various risks and the tools available to them to manage those risks.

Source: Modified based on Gentilini, Grosh, and Rutkowski 2019.

Essay 2: Where Does Targeting Fit Practically within Universal Social Protection?

Universal social protection is commonly conceived to be achieved by a system of programs meant to serve different risks and populations (Cecchini and Nieves Rico 2015; Jorgensen and Siegel 2019; Ortiz, Cummins, and Karunanethy 2017; UNRISD 2013). In their joint initiative on universal

social protection, the International Labour Organization and World Bank say the following:

> Universal social protection refers to the integrated set of policies and programs designed to provide income security and support to all people across the life cycle, with particular attention to the poor and the vulnerable. Universal social protection includes adequate cash transfers for all who need them, especially children; benefits/support for people of working age in case of maternity, disability, work injury or for those without jobs; and pensions for all older persons. This protection can be provided through social insurance, tax-funded social assistance/safety nets benefits, public works programmes and other schemes guaranteeing basic income security. Social protection programs aim at specific demographic groups of the population (e.g., children, persons with disabilities, women and men of the working age, older persons, etc.) and at households in chronic or transient (for instance, caused by shock such as a natural disaster) poverty. The objective of the social protection programmes is often not only to provide income support, but also to build up resilience to shocks and enhance connections to productive activities.[4]

In all countries, especially those with the highest coverage, social protection is built from a series of programs of different sorts—some programs pay out depending on the state of the worker (for example, for unemployment, retirement, illness, or disability) or individual (for example, age), and some pay out depending on the state of the family (for example, due to poverty or a natural disaster). Some programs pay small benefits to supplement the resources of families that are assumed to have at least some earned income (for example, for child allowances or some poverty benefits), and some pay larger benefits to substitute for income (for example, for unemployment, retirement, or disability). Some programs calibrate benefits to establish a minimum floor of well-being and some to compensate in whole or part for loss of income or assets. All these programs imply differentiation between groups and therefore require the administrative capacity to discern differences in needs.

Targeting is used in many aspects of social protection in countries that espouse the goal of universal social protection. The differentiation of eligibility or benefits is used to prioritize those most in need and/or fit benefit or services to purpose.

Universal social protection may require giving *subsidized insurance* to those with higher risk, worse outcomes, or lesser ability to self-insure. Subsidized health insurance has been an important means of moving toward universal health coverage. Cotlear et al. (2015) trace 24 countries seeking universal health care through bottom-up reforms. In several of these countries, the same social registries used to determine priority or eligibility for poverty-targeted programs were used to determine the

eligibility or degree of subsidization for subsidized health insurance or as a means to ensure outreach and enrollment in such programs. Packard et al. (2019) propose similar subsidies for pensions. For example, Chile provides subsidies to the solidarity fund portion of its unemployment insurance savings account (Sehnbruch and Carranza 2015).

Social pensions (pensions granted based on age, without the requirement of prior contribution) are another way to ensure that one of the functions of traditional social insurance is available, although without prior affiliation and collection. Nearly 100 countries have such programs, with various ways of combining them with contribution-based schemes—a score are universal for all individuals older than the defined age threshold; 28 countries grant social pensions only to those who do not receive contributory pensions; 55 use some type of welfare assessment or means test; and others are available only to those without (some minimum level of) contributory pension (HelpAge 2018).

Again, in something of an insurance role, governments often provide *responses to natural disasters*—by providing basic assistance in cash or in kind to maintain minimum living conditions in the immediate aftermath and/or providing assistance in rebuilding housing or the assets on which people's livelihoods depend (Bowen et al. 2020; UNICEF 2019). The population served may be those who suffer large losses, those whose losses drop them into a category of poverty, those who were already poor and are expected to become still poorer, or a combination of these groups. (Chapter 3 provides further discussion on this topic.)

Although broad-based or universal insurance programs and age-based programs can reduce some of the reasons that households fall into poverty, they cannot prevent it altogether. Thus, there is a need for programs of *last resort income support* for those who remain in poverty. A variant of these is the guaranteed minimum income programs that are common in Europe (Coady et al. 2021), which simultaneously target poverty and serve an insurance function (box 1.1). To work effectively as insurance, last resort income support must be funded as entitlements so that all eligible applicants receive benefits, and it must have excellent outreach and delivery systems so that all the eligible individuals apply and receive benefits.

Most countries have *programs for the poor or extreme poor* or a subset of them. If such programs are not surrounded by sufficient insurance or complementary programs, they may seem to be more "first resort" than "last resort" and therefore often have higher coverage than last resort income support. Most of the conditional cash transfer programs in Latin America and the Caribbean and East Asia fit this niche, at least at their outset, with coverage usually in the range of 10 to 25 percent of the population. These programs act as a single square in the patchwork quilt[5]

that builds toward the country's social protection system, or as a step along the path of the progressive realization of universal social protection (World Bank 2018b).

There is often a range of other *programs that aim to cover larger shares of the population* than just the poor, although not necessarily reaching universal coverage. Common examples are school lunch programs, "social tariffs" for utilities, and cash transfers in lieu of fuel subsidies, which may reach quite high into the income distribution. For example, Jordan's bread subsidy cash compensation reaches nearly 80 percent of households, and the (currently nonoperational) fuel subsidy cash compensation reached about 70 percent (Rodriguez and Wai-Poi 2020).

The COVID-19 stay-at-home orders and the ensuing economic crisis called forth a huge set of policy responses across the social protection sector. A common thread was to scale up support quickly and often quite broadly to reach groups well beyond the focus of the usual (often small) programs for the chronic poor. Coverage of large swaths of the "missed middle" or urban informal sector was desired, and countries innovated as best they could to approximate it. The Philippines initially planned to cover three-quarters of its population for two months under its emergency support program to support those most affected by enhanced community quarantine, with the coverage and length later extended in some areas and sectors.[6] Namibia rolled out its one-off Emergency Income Grant for working-age citizens who lost their jobs and were not receiving other forms of social protection, which eventually reached about 30 percent of the population (Gentilini et al. 2020, v12).

Occasionally, programs combine universality with selectivity, recognizing that the most disadvantaged need more, not equal, support. Various studies have put forth the notion of universal or high coverage with differentiated benefits as an option to be considered. Examples include the United Nations Children's Fund–Overseas Development Institute (UNICEF–ODI 2020) treatise in support of universal child benefits and Coady and Le's (2020) discussion of the role of universal and targeted programs in fiscal redistribution. Soares, Bartholo, and Guerreiro Osorio (2019) propose a country-specific program—a universal child allowance in Brazil. There are a few cases of such designs in current policy. Germany's child allowance covers all children up to age 18 but provides supplemental benefits to the needier children (ISSA 2018). In a cross-sectoral understanding of the same idea of helping those who are furthest behind, the Brazilian Bolsa Familia means-tested conditional cash transfer program was viewed by its designers as a means to help the poorest realize their rights to health and education (Campello and Neri 2013). India's Public Distribution System is a high-coverage but not universal program (covering about 75 percent of the rural population and 50 percent of the urban population). It provides the

opportunity for poorer households in the Antyodaya card holders category to purchase seven times as much subsidized food grain per month as the amount for other claimants (Drèze et al. 2018).

The share of programming that differentiates eligibility and benefits within the overall architecture of social protection varies greatly depending on the program mix, the shape of the economy, the degree to which the state is seen as provider of first resort or last resort, the tax structure, the maturity of the social protection system, and so forth. In general, the higher are the coverage and benefits of the insurance and universal programs, the smaller is the role of programs designed for only the poor or extreme poor. However, even in countries with a full range of programs, the role of poverty-targeted programs can vary. In the European Union, for example, the share of means-tested programming in the overall nonpension social protection sector varies from about 75 percent in the Netherlands and Portugal to less than 10 percent in the Baltics. And coverage of the poorest 40 percent is not as linked as might be expected. The Netherlands covers virtually all the poorest, while Portugal and Latvia cover about 30 and 35 percent, respectively, of the bottom two quintiles (Bussolo et al. 2018). In the Organisation for Economic Co-operation and Development (OECD) countries with universal child allowances, they account for an average of about 15 percent of the poverty reduction brought about by the transfer system (UNICEF–ODI 2020, figure 6).

The share and characteristics of the target population for each kind of program may vary. Eligibility levels can be set to cover only the extreme poor, encompass many who are not poor but may be at risk of poverty, or extend to the middle class. Indeed, sometimes thresholds are set so high that the term "affluence testing" is used, although the notion of granting or customizing benefits according to the welfare of the recipient is similar. Over the next decade, as countries work to accelerate progress toward universal social protection, governments may raise the eligibility thresholds for social assistance programs that are currently very narrowly focused, and they may broaden eligibility for subsidies for enrollment in insurance programs. The switch to programs covering relatively higher shares of the population may have implications for the choice of mechanisms to determine eligibility or customize benefits and the delivery systems that support them.

Essay 3: What Is the Rationale for Targeting by Welfare or Other Metrics?

Economics is "the science of scarcity"; thus, it focuses on how to get the maximum impact from any expenditure. In the traditional economic formulation, thinking about whether to focus resources on a particular group

can be posed as a problem of the efficient use of constrained resources (whether in mathematical formulae or more narrative or graphical form). Various authors have written about this topic over the years (see Besley and Kanbur [1990] for a much-cited formulation and Coady, Grosh, and Hoddinott [2004]; Coady and Le [2020]; and Devereaux et al. [2017] for more recent contributions).

The advantage of differentiating eligibility and benefits according to welfare is the power of focusing resources on those who need them most. For example, if reducing poverty is the objective, imagine an economy of 5 million people, of which 1 million are poor, each in need of an additional $30 per month to reach the poverty line. Poverty could be eliminated with a budget of $30 million per month if the funds could be perfectly directed to the poor with no costs. If, instead, funds were distributed equally across the whole population, it would take $150 million per month to eliminate poverty. Or if the budget remained at $30 million with funds distributed equally to the whole population, the poor would receive only $6 per month, much less than what is needed to get them to the poverty line. The power to have a bigger impact on poverty, or to reduce poverty at lower cost, is the main driver for differentiating benefits in social assistance programming.

Figure 1.1 provides a graphic generalization of the power to do more, or to make do with less. The figure is an adaptation of the expositions in Besley and Kanbur (1990) and Coady and Le (2020). This version assumes that it is simple to measure welfare and that there is a set poverty line[7] (topics taken up with more realism in chapters 3 and 6). People are arranged from poorest to richest along the x-axis, and with no policy, their pre- and post-transfer incomes are the same, so their welfare falls along the 45-degree line, AD. The figure shows the contrast between three policies:

- Perfect differentiation of benefits, as in a perfectly implemented guaranteed minimum income program, would give a transfer to each person who has an income less than Z, equal to the difference between the height of line Z and their individual income. For the poorest person, the transfer is AZ. For a person who is just at the threshold of poverty, shown at H, the benefit becomes zero. In this perfect targeting case, welfare would fall along ZHD. The budget of the transfers is the triangle AZH, shaded in dark blue. (This is a generalization of the case in which each poor person needed the same $30 to reach the poverty line, but it is analogous in that simplified example to paying $30 each to the 1 million poor and nothing to the nonpoor, with a budget of $30 million.)
- A universal benefit and the same budget as used in the perfect targeting scenario would result in welfare along line BC, where the area of budget

rectangle ABCD (shaded in light blue) equals that of triangle AZH. This would raise the welfare of the poorest person to B, and much of the budget would be spent on those above the poverty line. (This would be analogous to paying everyone $6 dollars, with a budget of $30 million.)

- A universal benefit that is large enough to eliminate poverty would result in welfare along line ZE, raise the income of the poorest person to the poverty/guarantee line, but use a budget of AZED, which is clearly larger than the budget in the prior scenarios. (This would be analogous to paying everyone $30, with a budget of $150 million.)

Figure 1.1, which illustrates the theoretical potential of targeting, relies on the assumption that it is possible and costless to identify who needs what benefits. But this assumption must be tempered by realism about errors and costs (themes that recur throughout the book, with their first detailed treatment in chapter 2).

- It is not possible to know perfectly who is poor or what their income is; errors will occur, often including substantial errors. The theoretical power of targeting will diminish as a consequence, by an amount that depends on the frequency of the errors and where they occur in the welfare distribution.

Figure 1.1   Contrasting Policy and Budget Scenarios: Base Case
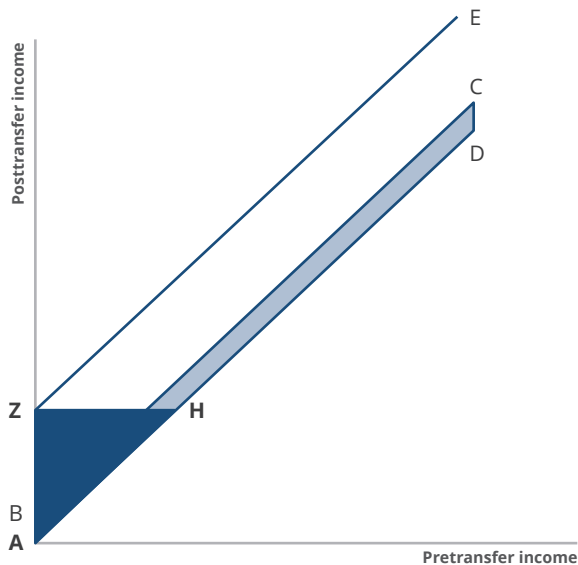
Sources: Based on Besley and Kanbur 1993; Coady and Le 2020.

- Moreover, such an attempt at selectivity or customizing benefits has costs:

  o There are administrative costs in gathering the information used to differentiate eligibility or benefits.
  o People may face transactions cost as part of the administrative process of proving their eligibility, which reduces the net benefit to them.[8]
  o There may be stigma or social discord as a result of making distinctions between people of different welfare levels.
  o Some criteria for determining eligibility or customizing benefits could create unwanted incentives. Particular attention is given to the concern that families could reduce work effort to stay poor enough to qualify for benefits in programs with an above/below threshold, or they may cease to work altogether with a minimum income guarantee formulation.
  o Political support for programs that treat some people differently than others may be less than for those that treat all alike. And with lower political support, the budgets for the programs may be lower.

Targeting is not an objective itself but a tool to be deployed on a selective basis. Policy makers must decide whether, how broadly or narrowly, and how to target a program based on their appreciation of the magnitude of the benefits of concentrating resources where they are most needed versus the magnitude of the various errors and costs and how these vary among different policy options.[9] This book aims to provide policy makers updated information they can use in their decision-making.

The problem setup is simple and intuitive, but finding answers or generating consensus around the preferable degree of targeting is much harder. The budget and administrative costs are measured in dollars, but stigma and political division are not. The errors of inclusion or exclusion will vary depending on the data or method of selection used. There may be a link via political economy between the available budget, narrowness of selection, selection method, and errors. The trade-offs among some of the dimensions are not well mapped and, where they are partially mapped, they are somewhat variable across contexts. But raising taxes is always hard and there are so many calls for government expenditures that the question of how to make the best use of scarce resources keeps the issue of targeting the neediest perennially on the table.

The remainder of the book helps in generalizing from or thinking about different parts of this simplified version of the pros and cons of differentiating eligibility or benefits to concentrate resources on those most in need.

Essay 4: How Does Thinking about Shocks Rather Than Static Poverty Change the Framework?

Shocks imply that some people lose income or assets.[10] Shocks that affect small numbers of unlinked people—such as an individual losing a job or a tree falling over and damaging the roof—are called idiosyncratic shocks. Shocks that affect many people at once—a whole industry reducing employment or a storm damaging buildings over a large geographic area— are called covariate shocks.

Public support may be offered to those whose welfare falls below the poverty line or an otherwise defined eligibility threshold and/or to those who have suffered losses above a certain size. Building on figure 1.1, figure 1.2 repeats the scenarios of perfect targeting and universal benefits with the same budget. Figure 1.2 omits the scenario of a universal basic income with a benefit equal to the poverty line since that is not a policy option that has been widely practiced (Gentilini et al. 2019). The figure adds the much more typical program design of a flat benefit to all who fall below the eligibility threshold, with a smaller budget of ABFH. The different colors represent different individuals—Red, Yellow, and Green—at different levels

Figure 1.2 Contrasting Policy and Budget Scenarios: With Shocks
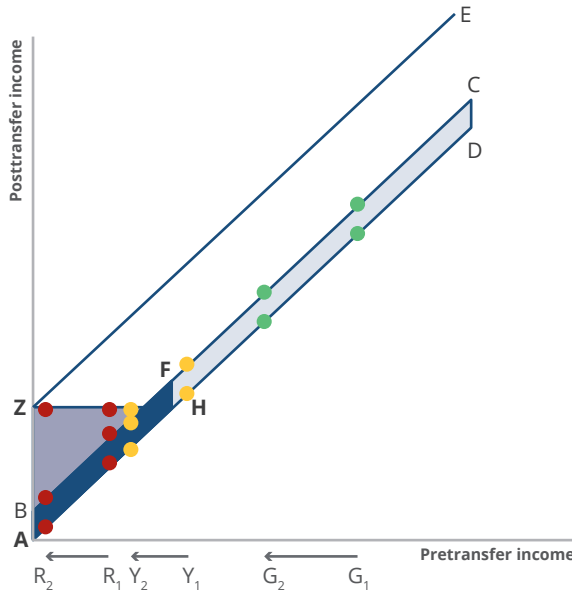
Sources: Based on Besley and Kanbur 1993; Coady and Le 2020.

of welfare. Red's pretransfer income is the lowest of the red dots at R_1, with the posttransfer income depending on which policy option is chosen, and similarly for Yellow, Y_1, and Green, G_1. A shock lowers the incomes of all three individuals (whose positions shift left, as indicated by the arrows) between periods 1 and 2. Red's initial income was below the poverty line, and it falls due to the shock. Yellow, who was initially not poor, has income that falls below the poverty line as a result of the shock. (In this example, Red and Yellow have a similar size loss of income.) Green remains above the poverty line but suffers a much larger shock in income. Depending on the blend of objectives between poverty reduction and risk management and the specific design parameters for how much differentiation there is in eligibility and benefits, different policy scenarios are possible. With a guaranteed minimum income program, Red would receive greater assistance in period 2 because Red is poorer (the distance between Red's dots and line ZH is greater). Yellow could start to receive assistance but would get only a modest payment as the "after" dot is not far below guaranteed income line ZH. With simpler programming, say with eligibility but not benefits differentiated by welfare, Red would not receive a larger payment, but Yellow would start to receive a payment. With a focus only on poverty, Green would receive nothing. With a focus only on loss, Red, Yellow, and Green would receive something. With the benefit related to the size of the loss, Green would get a larger payment than Red or Yellow, although Green would remain the most well-off among the three. With a benefit threshold only for very large losses, only Green would benefit.

All the barriers to "perfect targeting" in a static scenario carry through to a shock scenario, and some have greater force. Measuring income or assets remains difficult. If the objective is to measure losses, it implies that measures are needed for before and after, which not only implies two points of data, but also that errors will be compounded. Political economy and stigma must be considered, although support for action may be greater and stigma may be lower if suffering the shock could happen to anyone, unrelated to their effort or degree of social inclusion/exclusion. But the issue of overall program costs carries through as well—to restitute all losses is costly, possibly enormously costly, and so the question of whether to differentiate support by welfare or losses remains.

Helping to manage shocks has been such an important area for social protection programming in recent years that the term "adaptive social protection" has been coined to draw attention to the need to ensure that individual programs and the mix of programs are fit for the task. The term adaptive social protection gained traction around natural disasters, especially as climate change heightened concern about the toll that drought, storms, and extreme heat were taking on rural livelihoods and the poor more generally. But the term and such considerations

generalize to other shocks, including economic shocks, whether occasioned by a global financial crisis or global pandemic. And it recognizes that helping households manage idiosyncratic shocks is an important aspect of social protection.

Adaptive social protection includes building resilience ex ante and responding quickly and appropriately once shocks hit, and both may imply targeting. Because the poor or near poor have so little ability to handle losses, improving the level or reliability of their incomes is important. Thus, programs that are designed to build resilience through training, financial inclusion, and asset transfers often focus on the poor, differentiating eligibility. The package of services is often common to all participants, although sometimes the size of the asset transfers may vary as well as whether they are provided as grants or loans. To ensure programs' ability to respond with agility when a shock hits, systems and financing need to be set up ex ante as well, although responses are likely to be differentiated at least somewhat by the severity of the calamity. Natural disaster responses usually focus on geographic areas. Responses to economic crises may focus on the poorer, or on those whose jobs are the most affected. Responses to idiosyncratic shocks may require fewer resources overall, but an agile mechanism is needed. How delivery systems and targeting mechanisms can serve or be adapted to serve in response to shocks is a recurring theme in chapters 3, 4, and 5.

Essay 5: Is Targeting the Poor Important for Outcomes Other Than Poverty?

Larger effects on human capital and economic behaviors among the poorest are consistent with the logic that the marginal impact of a dollar of transfer income declines with base income. The logic suggests that the impact of a $1 transfer on a person living on $1 a day is much greater than that on a person living on $5, $10, or $50 a day. This gives additional weight to the economic rationale for focusing resources on the neediest—on whom the impacts on poverty and other dimensions of welfare will be greatest. This theory has been confirmed by evidence.

The body of impact evaluations confirms that social assistance transfers reduce immediate money metric poverty, but also that they improve a long vector of outcomes that are commonly associated with poverty or viewed as part of multidimensional poverty. For example, research on this looks at cash transfers (Attah et al. 2016; Bastagli et al. 2016; Davis et al. 2016), public works programs (McCord and Slater 2009; Subbarao et al. 2013), school feeding programs (Bundy 2011; Drake et al. 2017), Africa (Ralston, Andrews, and Hsiao 2017), and Asia (World Bank and DFAT, forthcoming).

Research shows that to various degrees, depending on program design and context, transfers have raised school enrollment; increased the use of health services for children; improved the mental health, happiness, or optimism of family members; raised social capital; reduced intimate partner violence; reduced risky behaviors among teens; allowed households to invest in their livelihoods, pay down debts, or save; and generated positive local multiplier effects. Research shows that some programs have increased nutritional outcomes or learning. This evidence draws from a large range of programming with various targeting methods and outcomes. Importantly, where study design has allowed such measurement, the impacts have been larger among the groups within the program that were most disadvantaged at the beginning (Bastagli et al. 2016; OECD 2019). For example, in looking at the protective effects of Ethiopia's rural Productive Safety Net Program in the months since COVID-19 hit, Abay et al. (2020) show that the program generally protected food security among participant households, with food security indicators for participant households declining much less than those among the poor, nonparticipant households. Moreover, the program had the greatest effects among the poorest participants and those in the most remote areas. In an evaluation of a school feeding program in Ghana, no effects were found on height for age for the whole treated population (all income levels and ages 5–15). However, disaggregating the results by poverty status highlighted a positive effect of school meals on height for age in children ages 5–8 in poor households of 0.21 standard deviation, nearly twice the size of the effect observed in the overall population ages 5–8 (Gelli et al. 2019).

In their evaluation of the Pakistan Waseela-e-Taleem conditional top-up to the Benazir Income Support Program, Cheema et al. (2016) find that the impact of the Waseela-e-Taleem program on enrollment was higher for children in the poorest third of households in the evaluation sample, at 18 percentage points, compared with 8 percentage points for children in the other two-thirds of households. In Cambodia, Filmer and Schady (2008) show that the impact of the Japan Fund for Poverty Reduction scholarship program on enrollment was approximately 50 percentage points for girls in the poorest two deciles of a composite measure of socioeconomic status, compared with 15 percentage points for girls in the richest two deciles. For Nicaragua, Maluccio and Flores (2005) show marked differences in impacts on school enrollment by initial welfare levels—about 5 percentage points for the nonpoor, 15 percentage points for the poor, and 25 percentage points for the extreme poor. For Indonesia, Sparrow (2004) shows that a scholarship program implemented during the East Asia financial crisis had the largest effects on the poorest students. And impacts are often larger when the transfer is higher (for example, Bastagli et al. 2016; Ralston, Andrews, and Hsiao 2017).

Thus, equity and efficiency objectives are aligned. Focusing resources on the poor(er) helps to improve equity. It also helps to improve human capital efficiently.

Essay 6: Why Is Redistribution Important?

Distribution has intrinsic importance, and some sense of fairness is a basic part of the social contract, especially when individuals can do little about the sources of inequality. The intrinsic value of fairness and equity is so well accepted that arguments to support it are not repeated here (World Bank 2006, chapter 4). The breadth of support is crystallized in Sustainable Development Goal 10—to reduce inequality among and within nations. This could be called the intrinsic case for concern about inequality. There are also instrumental reasons for redistribution.

A first instrumental reason to focus on inequality is because at normal or likely rates, growth alone in many places is insufficient to reach poverty reduction goals. Calculations showing how different combinations of growth and inequality will affect poverty have a long history and can be done at the aggregate or national level (for example, ECLAC [2002] for Latin America; Yemtsov et al. [2019] for the Russian Federation). Moreover, as countries become less poor, inequality-reducing policies are likely to become relatively more effective for poverty reduction than growth-promoting policies (Olinto, Lara Ibarra, and Saavedra-Chanduvi 2014). The World Bank's pre-COVID-19 flagship report on Poverty and Shared Prosperity calculated that even if the world had grown at twice its historical rate, it would not have been enough to meet the goal of 3 percent extreme poverty ($1.90/day) by 2030 (World Bank 2018a). Much more pro-poor growth or redistribution would have been needed to bring the goal into sight.

Post-COVID-19, the challenges for poverty reduction and inequality loom larger. The World Bank's post-COVID-19 flagship report on Poverty and Shared Prosperity estimates the biggest increase in poverty in decades and expects a significant increase in inequality, although it does not produce a headline number (World Bank 2020b). The COVID-19 shock caused the greatest learning losses in poorer countries and among poorer children, and higher reductions in work among low-wage earners, youth, and women (IMF 2021; World Bank 2021b). Without strong interventions, the crisis may trigger cycles of higher income inequality, lower social mobility among the vulnerable, and lower resilience to future shocks. IMF (2021) demonstrates that inequality is rising in many advanced and large middle-income countries, comprising about two-thirds of the world's population, although it is falling from high rates in some others.

A second instrumental reason to focus on reducing inequality is because of the toll it can take on growth. Ostry, Berg, and Tsangarides (2014) review the growing literature on the theme and extend it, showing that lower net inequality is robustly correlated with faster and more durable growth, for a given level of redistribution. They show that it appears that redistribution has a generally benign impact on growth; only in extreme cases is there some evidence that redistribution may have direct negative effects on growth. Dabla-Norris et al. (2015) further support the case for redistribution, such as that achieved by progressive tax-transfer policies. They find that if the income share of the top 20 percent increases by 1 percentage point, gross domestic product (GDP) growth is actually 0.08 percentage point lower in the following five years, suggesting that the benefits do not trickle down. In contrast, a similar increase in the income share of the bottom 20 percent (the poor) is associated with 0.38 percentage point higher growth.

While a broad spectrum of structural policies can help reduce inequality in the long run—competition policies; equity in the provision of water, sanitation, transport, power, communications services, education, and health care; minimum wage and other labor market regulations; and so forth—the tax-transfer system has a more direct effect, with impacts possible in a short period rather than gradually over decades. Therefore, social protection programs are an important part of the toolkit to build equitable societies. Moreover, income inequality is linked to inequality of opportunity (IMF 2021), so well-focused, short-run actions to reduce income inequality can support structural efforts as well.

Essay 7: What Does the Distribution of Taxes Imply about the Distribution of Transfers?

Because the objectives of social protection include reducing poverty and/or inequality, the distribution of transfers and the taxes that support them must be considered together. Governments can achieve redistribution with flat (uniform) benefits if the taxes that support them collect (absolutely) more from those with higher welfare than those with lower welfare.[11] To achieve a given level of redistribution, the more sharply progressive taxes are, the flatter may be the benefits and vice versa. Although this point is conceptually obvious, because data have been scarce, joint consideration of the empirics of tax and transfer systems was not done as a matter of course.

In advanced economies, direct taxes and transfers reduce income inequality on average by about one-third (from a Gini of 0.49 to 0.31), with three-quarters of this reduction achieved through transfers,[12] which reduce inequality more at the bottom while taxes do so more at the top. IMF (2017)

notes that in recent years in some advanced economies, redistributive efforts have lessened despite increased inequality in labor incomes, confirming a trend noted by Bastagli, Coady, and Gupta (2012). The lessening of progressive taxation is a theme picked up in Bussolo et al.'s (2018) study of the social contract in Europe (both Western and Eastern). They show that from the early 1980s until prior to the global recession, the share of top incomes grew, while the top personal income tax and corporate tax rates fell sharply. In the United States, tax rates on the highest income earners (95th percentile and above) fell enormously and more or less steadily from 1950 to 2018, while remaining relatively flat and constant for the rest of the distribution. By 2018, tax rates were relatively flat overall, and indeed slightly lower for the very top earners than for the lowest decile (Saez and Zucman 2019). In the OECD, the top rates for personal income taxes, dividend income, interest earnings, and corporate income have each fallen on the order of 20 percentage points (IMF 2021).

In developing economies, fiscal redistribution is much more limited, reflecting lower revenues as well as a less progressive taxation mix (Fuchs, Sosa, and Wai-Poi 2021). Total tax revenues determine how much public spending, including on redistributive policies, can be done. As figure 1.3 shows, advanced economies raise tax revenues equivalent to around 25 percent of GDP. This falls significantly for developing countries, with non-high-income countries averaging considerably less than 20 percent. Moreover, within these lower levels of tax revenues, the mix of taxes on which developing countries rely is often less progressive than in advanced economies, where over a third of all revenue is from personal income taxes, which are the most progressive, and over half is from other progressive direct taxes, such as property, payroll, and corporate income taxes. Instead, indirect taxes on consumption (such as value added and excise taxes) make up the majority of the tax revenue base in developing countries as well as non-OECD high-income countries. Such taxes, which usually impose the same rate for all purchasers of goods, are more regressive than income taxes.[13] Even when exemptions and lower rates are placed on staples consumed more by the poor, richer households enjoy more of the savings due to their higher consumption.

Evidence for jointly considering the tax and transfer impacts in developing countries has been boosted by the Commitment to Equity project in recent years, as well as its inclusion as a Sustainable Development Goal indicator in March 2020, and shows a variable and cautionary picture. Inchauste and Lustig (2017) initially compiled comparative evidence from 17 low- and middle-income countries. For the years studied, the effect of taxes and transfers on the Gini was limited, at less than 4 points in all countries, with only South Africa (7.7) as an exception.[14] Rodriguez and Wai-Poi (2020) include a wider comparison of fiscal redistribution in 42 developing

Figure 1.3   **Composition of Tax Revenues, by Income Category (percent of GDP)**
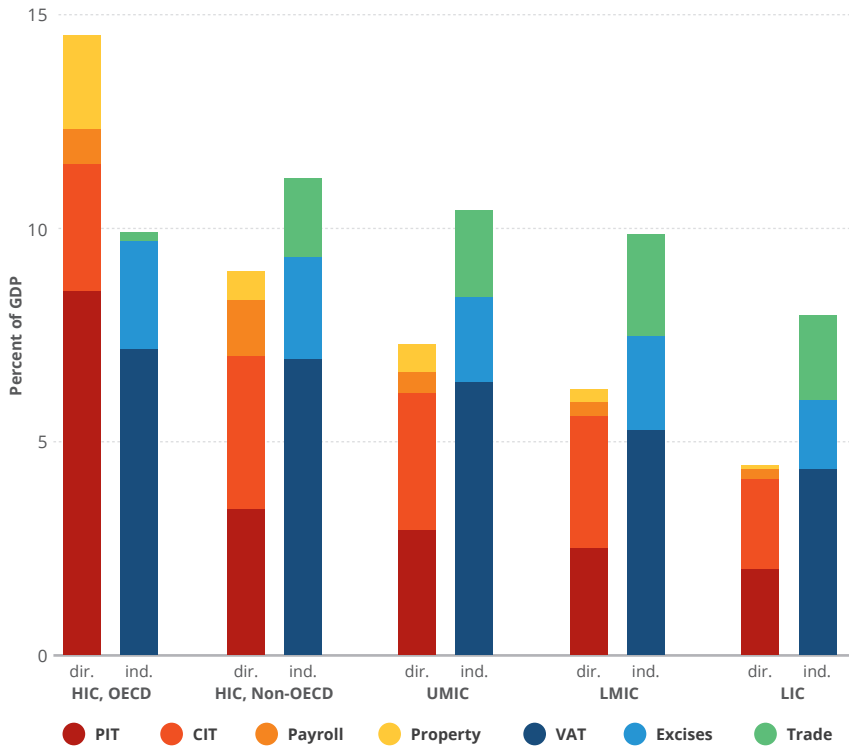
Sources: Fuchs, Sosa, and Wai-Poi 2021; calculations based on data from the International Monetary Fund's World Revenue Longitudinal Data set (WoRLD).

Note: Direct taxes (dir.) include PIT, CIT, payroll tax, and property tax. Indirect taxes (ind.) include VAT, excise tax, and trade tax. CIT = corporate income tax; GDP = gross domestic product; HIC = high income countries; LIC = low income countries; LMIC = lower-middle-income countries; OECD = Organisation for Economic Co-operation and Development; PIT = personal income tax; UMIC = upper-middle-income countries; VAT = value added tax.

countries; only 7 countries see a reduction in inequality of 5 points or more, with an average reduction of only 2.8 points (figure 1.4). In contrast, the average for the European Union is a difference of 21 Gini points between market and disposable income (Bussolo et al. 2018).[15]

With the focus on poverty rather than inequality, the results of the Commitment to Equity project become even more negative. Taxes and transfers reduce the poverty headcount in only half the countries studied by Inchauste and Lustig (2017). In the other half, poverty increases after direct and indirect taxes, even after taking into account the benefits of

Figure 1.4 **Fiscal Impact of Monetary Inequality (points reduction in Gini Index)**
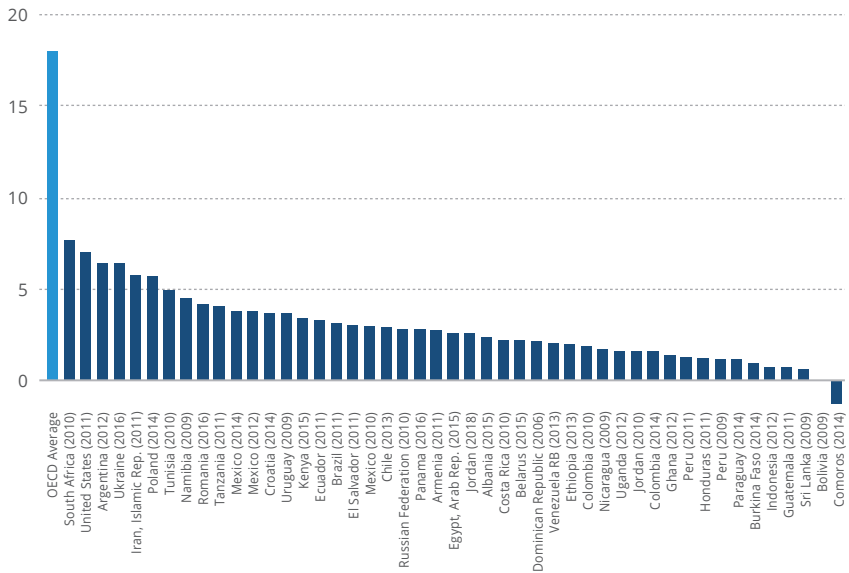
Sources: Based on Rodriguez and Wai-Poi 2020 based on the Commitment to Equity Institute database; 2018 results for Jordan; IMF 2017 for the OECD average.

Note: Developing country results exclude the impact of in-kind health and education expenditures but include indirect subsidies and taxes. OECD results exclude both in-kind spending and indirect subsidies and taxes. OECD = Organisation for Economic Co-operation and Development.

direct transfers. This happens in Armenia, Bolivia, Brazil, Ethiopia, Ghana, Guatemala, and Sri Lanka.

As countries develop and refine their social protection systems, they need to consider the implications of their tax system for the degree of targeting they want in their expenditures. Governments can achieve redistribution with flat (uniform) benefits if the taxes that support them collect more from those with higher welfare than those with lower welfare. Countries that generate more revenue can afford more complete social protection systems and have more leeway to choose universal or affluence-tested designs over more narrowly focused ones for a wider range of programming. In countries with particularly low expenditure on social protection, there is a particular need to focus on raising resources—even the power of targeting cannot make up for gross insufficiency of funding.

As developing countries look to expand their total tax revenue, they will need to consider the distributional implications. The April 2021 *Fiscal Monitor* (IMF 2021, 17) signals that progressivity and revenue

performance could be improved through broader tax bases; more progressive personal income taxation; more neutral capital taxation; improvements in the design of value added taxes; more and better use of carbon, property, and inheritance taxes; digital enhancements; and institutional strengthening to enable revenue administrations to implement and manage these tax reforms.

Essay 8: Can Budgets Be Raised over Time to Reduce the Need for Targeting?

One of the main reasons to target is to focus limited resources on the neediest, for example, to reduce poverty at low cost. Many targeted programs got their start in moments of crisis—following the debt crisis of the 1980s; the East Asia financial crisis of 1998; the global food, fuel, and financial crisis in 2008–09; the COVID-19 pandemic and linked recession; or national rather than global economic crises, droughts, floods, or other disasters. In crisis, needs are higher than normal and fast action is imperative to prevent or minimize big upticks in poverty or losses in human capital. In crisis, there may be an urgency that redirects budgets to the new priority or overrides the usual fiscal caution, but there is no time to build new sources of revenue and budgets can be quite constrained relative to need. Thus, focusing resources as best as possible, even if rather imperfectly, is not a surprising choice in crisis-driven programs.

In less pressured moments, when the question is not so much "what can be done today" but "what kind of society or social contract is it desirable to build," the budget constraint may be taken as less fixed. Thus, many exercises in building a vision of social protection provide ideas for program design, coverage, and benefits that exceed current social protection budgets. Ortiz, Cummins, and Karunanethy (2017) provide a costing exercise for social floors for 57 low- and middle-income countries, and Durán-Valverde et al. (2020) contribute a post-COVID-19 update. Filgueira and Espindola (2015) offer a version of basic income transfers for children and the elderly in Latin America; ILO-UNICEF (2019) provides a vision for universal child grants; and Packard et al. (2019) describe a comprehensive program for social assistance-insurance. In all these exercises, the cost of the suggested programs is, in most countries, in excess of the 1.5 percent average currently spent on social assistance programs (World Bank 2018b), sometimes several times as large. For example, Durán-Valverde et al.'s (2020) estimates for child allowances, maternity, disability, and old-age benefits amount to 8.5 percent of GDP for low-income countries, 3.4 percent for lower-middle-income countries, and 3.2 percent for upper-middle-income countries.[16]

Given the resonance of the larger visions, there is also work internationally that thinks about how to build fiscal capacity for social protection through raising additional revenue or reallocating expenditures. For example, Ortiz, Cummins, and Karunanethy (2017) outline possible vectors for action: reallocating public expenditures, increasing tax revenue, expanding social security coverage and contributions, lobbying for aid and transfers, eliminating illicit financial flows, using fiscal and foreign exchange reserves, managing debt, and adopting a more accommodative macroeconomic framework. Hoy and Sumner (2003) calculate the marginal tax rates on those above the $10/day and $15/day income thresholds, and possible reallocations from energy subsidies and "excess" military spending that would be needed to end poverty at the $1.90/day and $2.50/day thresholds. They find that one or a combination of these means is sufficient to end three-quarters of global poverty.

Over the past decade, some governments have been finding ways to boost spending on social assistance. For example, a study of seven Latin American countries (Argentina, Brazil, Colombia, Ecuador, Mexico, Peru, and Uruguay), covering about 75 percent of the region's population, shows that social assistance spending rose from 0.43 to 1.26 percent of GDP from 2003 to 2015 (World Bank 2018b). The increase in social assistance spending accelerated around the time of the 2008 financial crisis, despite a reduction in the rate of economic growth. In Europe and Central Asia, tracking 15 countries that represent about 60 percent of the population shows a more moderate increase in spending. The analysis suggests that in this group of countries, average spending rose from 1.2 to 1.8 percent of GDP from 2003 to 2009, and then fell slightly, to 1.6 percent in 2014. Before the financial crisis, the region seems to have reached a steady level of social assistance spending, then spending grew in response to the financial crisis, and then it converged to the prior level. For other regions, comparable data sets were not available over long periods of time. However, several countries have rolled out and increased expenditure on significant flagship social assistance programs (World Bank 2018b).

Social assistance spending is mostly financed through general revenues, but there have been innovations in financing as well, or redirecting of inefficient existing spending.[17] For example, Brazil used 21 percent of a financial tax from 1997 to 2007 to finance social insurance, 21 percent to finance its conditional cash transfer program, and 16 percent for other social spending. By 2007, this tax accounted for 1.4 percent of the GDP, which was sufficient to cover the cost of the conditional cash transfer and other noncontributory social protection programs. At the same time, rural social pensions were financed by a 2.5 percent urban wage levy. Bolivia and Zambia used revenues and taxes from natural resources to fund old-age pensions and child grants. Mongolia did so as well from 2010 to 2016 and

since then has moved to general revenues. Other countries have used indirect taxes to finance social spending. By 2010, 40 percent of Argentina's pensions were directly financed by consumption taxes; 2.5 percent of value added taxes in Ghana pays for social health insurance; Algeria and Mauritius use tobacco taxes to supplement social security revenues; and sin taxes in the Philippines, accounting for 1 percent of GDP in 2015, financed the extension of subsidized health insurance to 40 percent of the population as well as insurance coverage for the elderly.

Table 1.1 summarizes the different revenue sources and their potential impacts. On the spending side, Indonesia reduced energy subsidy spending by 71 percent between 2014 and 2017 and redirected it in part to investments in infrastructure, health, and social protection, the latter increasing by 28 percent over the same period. The Islamic Republic of Iran (2010), Jordan (2013), and Pakistan also replaced energy subsidies with new or expanded cash transfers. Ultimately, a combination of both new revenues and redirected spending can be used; this has been the case for social pensions in Bolivia, Brazil, Costa Rica, Mexico, and Thailand. It was recently estimated that a combination of closing value added tax exemptions, further reducing energy subsidies, and increasing tobacco excises would

Table 1.1   Potential Sources of Revenue

Revenue Source	Revenue potential	Growth friendliness	Redistributive potential	Costs of administration	Cost of compliance
Personal income taxes	Variable	Low	High	High	Medium/High
Corporate income taxes	Medium	Low	Low	Medium/High	High
General consumption taxes	High	Medium	Low	Medium	Medium
Excises	Medium/Low	Medium	Low	Low	Low
Property taxes	Medium/Low	High	Medium/High	High	Medium
Social security contributions	Medium	Low	Low	Low	Low
Green taxes	Low	Medium/High	Medium/Low	High	Medium
User fees	Medium	Medium/High	Low	Medium	Medium/Low
Royalties	Medium/High	Low	Low	Medium	Medium

Source: World Bank and DFAT, forthcoming.

generate 1.8 percent of GDP in Indonesia, which would be more than adequate to fund the 1.5 percent of GDP increase needed for social protection.

Economic stimulus packages around COVID-19 in their first six months cost more than twice what was spent during the Great Recession of 2008–09. In their COVID-19 responses, low-income countries were constrained and mustered about 2.5 percent of GDP for stimulus, about half the average of approximately 5 percent of GDP in other emerging and developing economies, and far less than the 10+ percent average in advanced economies. About 18 percent of stimulus spending was devoted to social protection. Average COVID-19 emergency social protection spending was US$243 per capita—ranging from US$695 in high-income countries to only US$4 in low-income settings. The latter amounts to only 0.51 percent of GDP per capita. For low-income countries, the financing was all external. Domestically, lower- and upper-middle-income countries have financed 37 and 47 percent, respectively, of their policy responses. High-income countries have financed 100 percent of their response domestically. The common approach was restructuring or reprioritizing budget lines, but nearly half the countries incurred domestic debt and deficit spending, while others tapped state reserves, contingent funds, and fiscal savings (Almenfi et al. 2020).

Post-COVID-19, the fiscal prospects are among the grimmest seen for many years. The International Monetary Fund's January 2021 *Fiscal Monitor* shows the world average fiscal deficit at 11.8 percent of GDP, which is nearly 8 percentage points higher than pre-COVID-19, and world debt levels at 98 percent of GDP, up 15 percentage points. Even with growth expected to resume in 2021, for emerging and developing countries, aggregate output in 2022 is expected to remain 6 percent below its prepandemic projection, and the pandemic will leave lasting scars on productivity, including through its effects on the accumulation of physical and human capital (World Bank 2021a). Moreover, even before COVID-19, emerging market and developing countries faced a projected weakening of potential growth in the next decade. Their government debt had risen by 11 percentage points in the past decade, and their fiscal deficits had widened substantially after the 2008–09 crisis, peaking in 2016 (World Bank 2019). In Africa, for example, which accounted for just over half of world poverty in 2015, half the countries were at high risk of or in debt distress at the end of 2018 (Beegle and Christiaensen 2019). In International Development Association (IDA) countries, government debt increased by 15 percentage points of GDP between 2011 and 2019. Government debt-to-GDP ratios could rise by a further 8 percentage points in 2020. Moreover, in IDA countries in 2020, government revenues fell from 15.7 to 15.0 percent of GDP, reversing the progress made on domestic resource mobilization since 2012 (World Bank 2020a).

Social protection is not the only sector with high-priority needs. Gaspar et al. (2019) calculate that to cover the costs of the Sustainable Development Goals related to health, education, electricity, roads, water, and sanitation (but *not* including social protection) would take on average about 4 percentage points of GDP in emerging markets, but 15 percentage points of GDP in low-income countries. They consider that increasing the tax-to-GDP ratio by 5 percentage points of GDP in the next decade is an ambitious but reasonable target in many developing countries, leaving a large gap between plausible resources and high-priority needs. This suggests that there will be fierce competition for resources among high-priority expenditures and thus highlights the importance of the political economy that shapes those decisions.

The battle for fiscal space is not easy. It is rarely easy to raise taxes, and there are pressing needs for many good purposes. Thus, the question of whether to target to conserve resources or to raise more revenue to allow broader social protection programs is perennial.

Essay 9: Does Universality Increase Budgets and Thus Reduce the Need for Prioritizing the Needy?

In the discourse on the political economy of budgets, taxes, and targeting, "more for the poor is less for the poor" has become something of a mantra. An important source of support for the idea is the median voter theory, which postulates that voters will vote for programs that benefit them directly. Thus, a program for a minority such as "the poor" will garner little political support, while one that extends benefits to the middle class or universally will garner enough votes to have much larger budgets. The analytical underpinnings of the argument have been developed by serious scholars (Gelbach and Pritchett 1997; Meltzer and Richard 1981). Several country-specific explorations of some aspects of the theory support it. Jacques and Noel (2018) provide one of the supportive cross-country findings for OECD countries. Taylor-Gooby's (2005) study is an example of the single-country literature. Looking at public opinion in the United Kingdom, he finds that there is broader support for the universal National Health System than for targeted social schemes. The argument has gained currency among institutions that advocate for social protection, such as the International Labour Organization, the United Nations Children's Fund, HelpAge International, the Global Coalition for Social Protection Floors, and Development Pathways. The idea seems to be so widely accepted that this chapter does not include a full literature review of support (see UNICEF–ODI 2020).

The optimism that high coverage will yield many votes and thus large budgets is intuitive and appealing since it removes the budget constraint that is the impetus for targeting, but it may be that the catchy phrase over-sells the idea. Thus, the following paragraphs consider whether the simple idea and phrase reflect the complexities of the world and empirical evidence.

Empirical research has not found strong support for the median voter theory and its central implication about redistribution. Casual observation reveals that no country has legislated a fully tax-financed and ongoing uni-versal basic income program,[18] much less one sufficient to prevent poverty (Gentilini et al. 2019). Even fully universal (purely age-based) child allow-ances exist in only 21 countries (UNICEF–ODI 2020) and universal (purely age-based) social pensions in 19 countries (HelpAge 2018). Milanovic (2000) offers a cogent review of the cross-country literature to that date and its mostly unsupportive findings. Milanovic uses more specialized data that began to become available from the newly emerging detailed and har-monized household surveys. He finds some support for greater redistribu-tion where factor income inequality is higher but less support for the specifics of the mechanism of the median voter theory. Acemoglu et al. (2015) provide a detailed literature review and empirics on the expected relationship between democracy, greater redistribution, and lower inequal-ity and why the expected relationships may not be realized. A particularly useful treatment by Coady, D'Angelo, and Evans (2019), using EUROMOD data, finds that when fiscal progressivity in tax-transfer systems is higher (for example, spending is more targeted), fiscal effort (spending) is some-what lower but not to an extent that offsets the effects of the redistribution (the systems cost a bit less but deliver more to the poor).

Voters may support a program for reasons that go beyond the short-term self-interest at the crux of median voter theory, which would imply that support would be wider than coverage in any given year. It may be that voters will support not only programs that benefit themselves in the year of voting, but also programs that promise coverage applicable to others of concern to the voter, such as poorer relatives, neighbors, or co-workers. Or voters may support a program for which they may not qualify at present but might need in the future if they face misfortune.[19] Or they may support a social contract that includes some sort of destitution prevention but wish to pay as little as possible for it. Klemm and Mauro (2021) report that in the United States, those who lost employment, suffered from COVID-19, or personally knew someone who had are more likely to support progressive taxation.

Giuliano and Spilimbergo (2014) show that those who lived through an economic depression in youth/early adulthood favor redistribution more strongly than those who did not. Alesina and Guiliano (2009, 2011) note

that a spell of unemployment can increase support for redistribution, and being African American (and thus part of a structurally disadvantaged group) does so more. Costa-Font and Cowell (2015) and Haggard, Kaufman, and Long (2013) take a wider look at social identities and how they interact with immediate economic self-interest to shape attitudes toward redistribution. They find that identities other than income/class influence support for redistribution. And for a sample of 34 countries in the European Social Survey, Olivera (2015) notes that religiosity (irrespective of which religion) is associated with stronger preferences for redistribution. Nowack and Schoderer (2020) show a mild, positive association between the strength of egalitarian values held by voters across countries and the share of universal social policies within social policy constructs. Similarly, Alesina and Glaeser (2004) and Graham (2002) find that support for redistributive policy is higher in countries where more people believe that poverty is due to structural causes or "luck" than individual effort or "laziness." Definitions of "deservingness" seem to matter for the sorts of programs that may garner support. For example, van Oorschot (2006) and van Oorschot and Roosma (2015) note that there are some regularities across Europe where support for the old and sick/disabled is higher than that for the unemployed or migrants.

It is important to consider policy making processes. Very rarely are voters able to vote on each policy individually (direct democracy) as assumed by the median voter theory. Rather, voters elect candidates or parties that represent composites of positions on many issues (representative democracy). Moreover, even in established democracies, policy making processes are not as egalitarian as the one-person-one-vote process at the ballot box. Various sorts of political participation increase with income (see Karabarbounis [2011] for evidence from the World Value Surveys in advanced OECD countries, or Harms and Zink [2003]). And voters are not the only voices. As van Oorschot and Roosma (2015) explain, many relevant groups—such as politicians, policy makers, administrators, street-level bureaucrats, representatives of interest groups, and experts—have opinions and may shape policy directly or through mass media discussions and portrayals of different target groups.

Framings of electoral processes that do not assume direct democracy suggest a less arithmetic link between the breadth of coverage and voter support. Alessina and Glaeser (2004) posit that differences between Europe's more universal and generous welfare state and the United States' more narrowly focused and smaller one stem from different forms of democracy (proportional versus majoritarian and the role of checks and balances in the Constitution), economic history, and the degree of racial and ethnic homogeneity. Iversen and Soskice (2006) also model and test how redistribution is differentially supported by proportional versus

majoritarian systems of democracy, finding more redistribution in proportional representation systems. Hickey et al. (2018) anchor their multicountry studies of the political economy of social protection in Africa on the "political settlements" framework, where the settlements are both between various factions of the elite and between the elite and nonelite. The importance of electoral politics can be low, depending on its chance of bringing about regime change. In turn, elite commitment to social assistance may stem from its ability to bolster the legitimacy of the regime in delivering on promises—for example, with respect to inclusive growth or poverty reduction.

Other writings on the political economy of social protection programs focus not just on how narrow or broad their client population is but also on other framing or design factors. Hickey et al. (2018) discuss the role of ideas and the link between the framing or program design and world views—especially around concerns about "handouts" and "dependency," but also recognizing collective responsibility for categories such as the elderly and deserving poor, including working-age adults and their dependents in times of drought. World Bank and DFAT (forthcoming) remind that social protection can be viewed in a variety of ways—as charity/costs; as an economic investment in poverty reduction, human capital, or productivity; as a part of the social contract; or as a justiciable right. These different views presumably imply different terms of support among voters or elites. The study also examines how design features of targeting, conditionality, and modality of transfer can shape support. Bossuroy and Coudouel (2018) pick up a similar theme, stressing the two-way interaction between social policy and politics. They articulate how program design features can garner support—not just for eligibility, but around conditionalities, recertification processes, productivity focus, and grievance and redress mechanisms. Davis et al. (2016) show how impact evaluation, especially when national actors are closely involved, can bolster the credibility of programs, strengthen the case for social protection, and address concerns about dependency or undesirable use of funds. And de Janvry et al. (2005) show that in Brazil, mayors facing reelection had an electoral advantage if the implementation of the Bolsa Escola program had no publicized errors of inclusion and/or had established local accountability councils, but they were not penalized for errors of exclusion (which were not under direct mayoral control as the budget was rationed by the federal government).

Better understanding of the nuances and variations in the response function between policy design and budgets could be a major contribution to social policy formulation. Meanwhile, policy makers are finding their way through their locally pertinent political environment with different salience given to social protection overall and to different programs or their features within the overall social protection system; different degrees of

support for state action for redistribution and constructs of deservingness or reciprocity; different numbers and coalitions of political parties; and different details of electoral and budget processes.

Essay 10: How Do Human Rights Frameworks View Targeting?

Human rights are a widely accepted lens through which to view the instruments and outcomes of social policy. Most countries have signed the several international human rights treaties that reference social and economic rights. Moreover, all United Nations agencies have committed to mainstreaming human rights throughout the United Nations System.

In contrast to the economic lens that analyzes social protection in terms of investments, costs, and constraints, the human rights lens takes as a starting point that countries have voluntarily taken on obligations to provide for economic, social, and cultural rights and must now honor that pledge via social protection, among other actions (ILO 2021a, 2021b; Sepúlveda 2016, 2018; Sepúlveda and Nyst 2012; UNRISD 2013). The Committee on Economic, Social and Cultural Rights states that the realization of the right to social security implies that states should take measures to establish social protection systems under domestic law and ensure their sustainability, that benefits are adequate in amount and duration, and that the level of benefits and the form in which they are provided are in compliance with the principles of human dignity and nondiscrimination. In complying with the right to social security, states must ensure that social protection is equally available to all individuals and in this respect direct their attention to ensuring universal coverage; reasonable, proportionate, and transparent eligibility criteria; affordability and physical accessibility by beneficiaries; and participation in and information about the provision of benefits (Sepúlveda and Nyst 2012, 19).

In principle, human rights standards are not compromised by the use of targeted schemes as a form of prioritization of the most vulnerable and disadvantaged groups. However, in accordance with human rights standards, the methods must comply with the principle of nondiscrimination, which not only requires that all eligibility criteria must be objective, reasonable, and transparent, but also entails an obligation to prioritize the poorest of the poor and avoid stigmatizing beneficiaries. Targeted protection must be implemented with the intention of progressively providing universal coverage (Sepúlveda and Nyst 2012, 38). Targeting is admissible within a human rights perspective and may even be a needed tactic to focus "first on the especially disadvantaged and marginalized individuals and groups" (UNRISD 2013). But the issue does not stop with the decision to target—it

carries through to the selection and operation of the mechanism to determine eligibility and the delivery systems (Sepúlveda 2018).

Differentiating eligibility or benefits while being compliant with human rights standards is challenging. Indeed, the litany of criticisms from the human rights perspective of money metric or proxy methods of selection can seem almost preclusive of their use. Many of those criticisms are not unique to the human rights perspective—they are shared by those who think in more economistic or political frameworks as well. Surely, there is much in targeting practice to improve upon from any perspective. Thus, the subsequent chapters revisit human rights perspectives while moving into the topics of the choice of methods and their implementation.

Conclusion

Programs that target abound throughout the social protection landscape, in countries that are rich and poor, with limited or nearly universal social protection systems. This alone is reason enough to study them: to understand how such programs can contribute to universal social protection systems and how differentiating eligibility and benefits can be done judiciously.

One of the principal motivations for targeting concerns the budget and the efficiency of its use. For a given budget, such prioritization can produce more progress on poverty reduction, income smoothing, and other dimensions of welfare such as human capital and inequality.

Another reason for targeting is to make programs fit for purpose. A person who is not suffering from a disaster is not a high priority for a disaster response; conversely, a family that is benefiting from a universal child allowance will need additional assistance if a hurricane destroys their crops and home. This logic applies to the list of objectives and programs of all the branches of social protection. Needs differ in degree and in the timing and level of support required.

Fiscal space is always contested, with many needs and visions across many sectors on how to use more resources compared with usually limited consensus on whether or how to raise taxes to finance such resources. Of course, the decisions on how to raise and share revenues are political and so political economy must be considered, accounting for the complexities of the institutions of democracy and the definitions of voters' preferences. In many places, achieving universal social protection will require much larger resources than are currently dedicated to the sector, which raises to highest importance the interplay between program features and political support. This chapter's reading of the many competing needs suggests that fiscal space will remain an important constraint on social protection expenditure in the near term, which will keep the debates about whether and how

much to target one program or another on the table. It is therefore important to continue to learn from experience about how social protection can be improved.

Although many programs try to differentiate eligibility and benefits, doing so is difficult. It cannot be done without errors and costs. Therefore, the next chapter delves into the empirics to understand the magnitudes of the outcomes and costs of the targeted social assistance programs observed in recent social protection programming in emerging and developing countries. The subsequent chapters take up the processes and methods used to differentiate eligibility and benefits, to learn how they can be done well.

Notes

1. As box O.1 in the overview chapter explains, "welfare" can be defined in various ways. Chapters 2 and 3 take up that discussion in more detail. This chapter uses the term without full specificity because the basic concern of focusing resources on those most in need pertains irrespective of the definition. The chapter uses a measure of money metric welfare as the default interpretation, and eligibility thresholds that can fall anywhere in the range from focusing on the very poor to screening out only the wealthy, but mostly fall below the median level of income.
2. ILO (2012); https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100: 0::NO::P12100_ILO_CODE:R202.
3. The formulae also commonly contain elements of redistribution, such as minimum benefits.
4. https://www.ilo.org/global/topics/social-security/WCMS_378991/lang--en /index.htm.
5. The term "patchwork" is sometimes thought of as derogatory, but it is useful to understand wherein the insult lies. Literally, patchwork is a specific design for quilting patterns (among other popular traditional designs, such as the northern star or wedding ring). Quilting originated as a practical way to produce warmth from pieces of fabric that were each too small to make a good blanket on its own. In that sense, quilting was a way to handle a budget constraint. When executed well, it produced both warmth and beauty from limited resources. The patchwork design is the simplest to execute as the shapes are simple squares, which are sometimes large and usually uniformly sized. Thus, patchwork is often the first pattern a novice quilter learns. The derogatory use of the term refers not to the idea that the object is a quilt, nor to its potential warmth or beauty, but to the skill of the seamstress.
6. https://www.msn.com/en-ph/money/personalfinance/social-amelioration -program-how-to-qualify-and-how-much-can-you-receive/ar-BB12EtA5; https://www.dof.gov.ph/dof-says-covid-19-emergency-subsidy-largest -social-protection-program-in-phl-history/.
7. The notion of "the poverty line" in this formulation can be interpreted more flexibly as an eligibility threshold or suite of thresholds for different programs

that may be lower or higher or dispersed around whatever poverty line is used for analytic and poverty tracking purposes. This is consistent with how policy is actually made in most of the world and generalizes the discussion of targeting beyond the cost-minimizing way of reducing absolute poverty to the much wider question of how to move resources toward the "left side" of the welfare distribution, perhaps with a notion of being more progressive than the market income, the taxes that would support the transfer, or an alternative social policy option.

8. They may also face transactions costs of receiving benefits, for example, time or fees incurred in going to pay points or maintaining bank accounts into which benefits are paid. But these costs are not related to the determination of eligibility for the program.

9. In economic pedagogic presentations, the solution to the problem is sometimes referred to as "optimal" targeting to emphasize the notion that, in mathematical parlance, "optimum" may be different from "maximum" or "perfect" targeting. This also draws parallels to the usage in optimal tax theory. In common English, optimal targeting sounds close in meaning to maximum, so we fear it would connote a prejudice toward focusing on perfect targeting. However, most countries and programs implement much more moderate solutions.

10. In mathematical terms, shocks can be positive or negative, but in English, the tendency is not to call sudden increases in income "shocks" but "good fortune." Thus, public policy is not generally concerned with positive shocks, although by virtue of higher income, a person might face higher taxes or lose eligibility for benefits from differentiated programs if the changes are large or permanent enough to be observed.

11. On the transfer side, use of the term "flat" is often in an absolute sense, such as when everyone gets a benefit of $20. On the revenue side, there is a tendency to refer to rates. So a flat tax rate could collect 10 percent of income from everyone. For a person with $100 in income, the tax would collect $10. For a person with $300 in income, it would collect $30. Thus, on net, a flat benefit of $20 financed by a flat tax rate of 10 percent would redistribute income to the first person (who pays $10 in taxes and gets $20 in transfers) from the second (who pays $30 in taxes and gets $20 in transfers).

12. This analysis does not include the impact of indirect taxes on consumption, such as goods and services, value added, or excise taxes. Such taxes are generally neutral at best and often somewhat regressive in formalized advanced economies.

13. The regressivity of indirect taxes can be overstated in developing countries, and indirect taxes may occasionally even be progressive (although not as progressive as income taxes). This is because poorer households buy more from informal locations and so pay less tax than those buying at formal locations. Bachas, Gadenne, and Jensen (2020) summarize the informality Engel curve for 31 countries.

14. There is an unresolved methodological question in the field. That is, should contributory pensions be treated as deferred income or transfers? The Commitment to Equity project presents all results with both calculations. This essay reports the results for contributory pensions treated as deferred income,

to focus more squarely on the noncontributory part of the social protection system where the debate over universal versus targeted benefits is most heated. If pensions are treated as a transfer, Georgia's fiscal system would reduce the pre- and postfiscal Gini's by 11.2 points.

15. The EU redistribution effect does not include indirect consumption taxes and indirect price subsidies (such as subsidies for food and energy, which are important in many developing countries). However, the average reduction in equality from prefiscal (market) income to consumable income (accounting for indirect taxes and subsidies) is also 2.8 points, so the discrepancy in fiscal redistribution between high-income countries and developing countries remains.

16. The analysis considers children between ages 0 and 5 years. The maternity benefit is for women ages 15 to 49 with newborns, and the numbers of beneficiaries are calculated based on the observed country-specific fertility rates. For disability benefits, the study only considers persons with a severe disability, on the assumption that participation in employment may be challenging and may require specific support such as transportation allowances; the size of the eligible population is obtained from country-specific disability estimates from the World Health Organization's database on estimated years living with disability. For old age, the potential beneficiary population includes persons ages 65 years and older. For children, the benefit is defined as 25 percent of the national poverty line. For maternity, the cash benefit is set at 100 percent of the national poverty line during four months around childbirth to protect the critical period when mothers and newborns are most vulnerable. For disability and old-age pensions, the amount of the benefit is 100 percent of the national poverty line.

17. See Barrientos and Lloyd-Sherlock (2017); Holmemo et al. (2020); HelpAge International (2011); ILO (2018); IMF (2019); Packard et al. (2019); and World Bank (2016) for further discussion of the examples in this essay.

18. Mongolia and the Islamic Republic of Iran came close with temporary programs (see Gentilini et al. 2019) and a handful of countries and economies, including Hong Kong SAR, China; Japan; the Republic of Korea; Serbia; Singapore; and Tuvalu, initiated universal temporary COVID-19 response programs (Gentilini et al. 2020).

19. In the well-known US food stamps program (formally the Supplemental Nutrition Assistance Program, or SNAP), caseloads are about 14 percent of the population in a given month. Over their lifetime, half of all US children have received support, implying that political support could be drawn from a much larger share of the population than the current caseload (Oliviera et al. 2018).

References

Abay, Kibrom A., Guush Berhane, John Hoddinott, and Kibrom Tafere. 2020. "COVID-19 and Food Security in Ethiopia: Do Social Protection Programs Protect?" Policy Research Working Paper 9475, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34794.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2015. "Democracy, Redistribution, and Inequality." In *Handbook of Income Distribution*, volume 2B, edited by Anthony B. Atkinson and François Bourguignon, 1885–966. Amsterdam: Elsevier. https://economics.mit.edu/files/10394.

Alesina, Alberto, and Paola Giuliano. 2009. "Preferences for Redistribution." Working Paper No. 14825, National Bureau of Economic Research, Cambridge, MA. https://www.nber.org/system/files/working_papers/w14825/w14825.pdf.

Alesina, Alberto, and Paola Giuliano. 2011. "Preferences for Redistribution." In *Handbook of Social Economics*, volume 1A, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 93–131. Amsterdam: Elsevier.

Alessina, Alberto, and Edward Glaeser. 2004. *Fighting Poverty in the US and Europe: A World of Difference*. Oxford: Oxford University Press.

Almenfi, Mohamed, Melvin Breton, Pamela Dale, Ugo Gentilini, Alexendar Pick, and Dominic Richardson. 2020. "Where Is the Money Coming From? Ten Stylized Facts on Financing Social Protection Responses to COVID-19." Social Protection and Jobs Policy and Technical Note No. 23, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34802.

Attah, Ramlatu, Valentina Barca, Andrew Kardan, Ian MacAuslan, Fred Merttens, and Luca Pellerano. 2016. "Can Social Protection Affect Psychosocial Wellbeing and Why Does This Matter? Lessons from Cash Transfers in Sub-Saharan Africa." *Journal of Development Studies* 52 (8): 1115–31. https://www.tandfonline.com/doi/full/10.1080/00220388.2015.1134777.

Bachas, Pierre, Lucie Gadenne, and Anders Jensen. 2020. "Informality, Consumption Taxes and Redistribution." Policy Research Working Paper 9267, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/33851.

Barrientos, Armando, and Peter Lloyd-Sherlock. 2009. "Pensions, Poverty and Wellbeing: The Impact of Pensions in South Africa and Brazil, a Comparative Study." Pension Watch, HelpAge International, London. https://www.helpage.org/silo/files/pensions-poverty-and-wellbeing-the-impact-of-pensions-in-south-africa-and-brazil.pdf.

Bastagli, Francesca, David Coady, and Sanjeev Gupta. 2012. "Income Inequality and Fiscal Policy." IMF Staff Discussion Note SDN/12/08, International Monetary Fund, Washington, DC. https://www.imf.org/external/pubs/ft/sdn/2012/sdn1208.pdf.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano. 2016. "Cash Transfers: What Does the Evidence Say? A Rigorous Review of Programme Impact and of the Role of Design and Implementation Features." Overseas Development Institute, London. https://odi.org/en/publications/cash-transfers-what-does-the-evidence-say-a-rigorous-review-of-impacts-and-the-role-of-design-and-implementation-features.

Beegle, Kathleen, and Luc Christiaensen. 2019. *Accelerating Poverty Reduction in Africa*. Washington, DC: World Bank.

Besley, Timothy, and Ravi Kanbur. 1990. "The Principles of Targeting." DEC Working Paper 385, World Bank, Washington, DC. https://imagebank2.worldbank.org/search/700331.

Besley, Timothy, and Ravi Kanbur. 1993. "The Principles of Targeting." In *Including the Poor*, edited by Michael Lipton and Jacques van der Gaag, 67–90. Washington, DC: World Bank.

Bossuroy, Thomas, and Aline Coudouel. 2018. "Recognizing and Leveraging Politics to Expand and Sustain Social Safety Nets." In *Realizing the Full Potential of Social Safety Nets in Africa*, edited by Kathleen Beegle, Aline Coudouel, and Emma Monsalve. Washington, DC: World Bank. http://hdl.handle.net/10986/29789.

Bowen, Thomas, Carlo del Ninno, Colin Andrews, Sarah Coll-Black, Ugo Gentilini, Kelly Johnson, Yasuhiro Kawasoe Adea Kryeziu, Barry Maher, and Asha Williams. 2020. *Adaptive Social Protection: Building Resilience to Shocks*. International Development in Focus. Washington, DC: World Bank. https://openknowledge .worldbank.org/handle/10986/33785.

Bundy, Donald. 2011. *Rethinking School Health: A Key Component of Education for All*. Washington, DC: World Bank. https://openknowledge.worldbank.org handle/10986/2267.

Bussolo, Maurizio, Maria Dávalos, Vito Peragine, and Ramya Sundaram. 2018. *Toward a New Social Contract: Taking on Distributional Tensions in Europe and Central Asia*. Washington, DC: World Bank. http://hdl.handle.net/10986/30393.

Campello, Tereza, and Marcelo Côrtes Neri. 2013. "Programa Bolsa Família: uma década de inclusão e cidadania." Organizadores: Tereza Campello and Marcelo Côrtes Neri. Brasília: Ipea. ISBN 978-85-7811-186-1.

Cecchini, Simone, and Maria Nieves Rico. 2015. "The Rights-Based Approach to Social Protection." In *Toward Universal Social Protection: Latin American Pathways and Policy Tools*, edited by Simone Cecchini, Fernando Filgueira, Rodrigo Martinez, and Cecilia Rossel. Santiago, Chile: Economic Commission for Latin America and the Caribbean. https://www.cepal.org/en/publications/39484-towards -universal-social-protection-latin-american-pathways-and-policy-tools.

Cheema, Iftikhar, Martina Garcia Asia, Simon Hunt, Sarah Javeed, Tanya Lone, and Sean O'Leary. 2016. "Benazir Income Support Programme: Evaluation of the Waseela-e-Taleem Conditional Cash Transfer." Oxford Policy Management, Oxford, United Kingdom.

Coady, David P., Devin D'Angelo, and Brooks Evans. 2019. "Fiscal Redistribution and Social Welfare." IMF Working Paper No. 19/51, International Monetary Fund, Washington, DC. https://www.imf.org/en/Publications/WP/Issues /2019/03/08/Fiscal-Redistribution-and-Social-Welfare-46588.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/14902.

Coady, David, Samir Jahan, Riki Matsumoto, and Baoping Shang. 2021. "Guaranteed Minimum Income Schemes in Europe: Landscape and Design." IMF Working Paper 2021/179, International Monetary Fund, Washington, DC. https://www .imf.org/en/Publications/WP/Issues/2021/07/02/Guaranteed-Minimum -Income-Schemes-in-Europe-Landscape-and-Design-461341.

Coady, David, and Nghia-Piotr Le. 2020. "Designing Fiscal Redistribution: The Role of Universal and Targeted Transfers." Working Paper 2020/105, International Monetary Fund, Washington, DC. https://www.elibrary.imf.org/view/journals /001/2020/105/article-A001-en.xml.

Costa-Font, Joan, and Frank Cowell. 2015. "Social Identity and Redistributive Preferences: A Survey." *Journal of Economic Surveys* 29 (2): 357–74. https://online library.wiley.com/doi/full/10.1111/joes.12061.

Cotlear, Daniel, Somil Nagpal, Owen Smith, Ajay Tandon, and Rafael Cortez. 2015. *Going Universal: How 24 Developing Countries Are Implementing Universal Health Coverage Reforms from the Bottom Up*. Washington, DC: World Bank. http://hdl .handle.net/10986/22011.

Dabla-Norris, Era, Kalpana Kochhar, Nujin Suphaphiphat, Frantisek Ricka, and Evridiki Tsounta. 2015. "Causes and Consequences of Income Inequality: A Global Perspective." Staff Discussion Note 15/13, International Monetary Fund, Washington, DC. https://www.imf.org/external/pubs/ft/sdn/2015/sdn1513 .pdf.

Davis, Benjamin, Sudhanshu Handa, Nicola Hypher, Natalia Winder Rossi, Paul Winters, and Jennifer Yablonski. 2016. *From Evidence to Action: The Story of Cash Transfers and Impact Evaluation in Sub-Saharan Africa*. Rome: Food and Agriculture Organization. https://oxford.universitypressscholarship.com/view/10.1093 /acprof:oso/9780198769446.001.0001/acprof-9780198769446.

de Janvry, Alain, Frederico Finan, Elizabeth Sadoulet, Donald Nelson, Kathy Lindert, Bénédicte de la Brière, and Peter Lanjouw. 2005. "Brazil's Bolsa Escola Program: The Role of Local Governance in Decentralized Implementation." Social Protection Discussion Paper Series 0542, World Bank, Washington, DC. http://hdl.handle.net/10986/20383.

Devereaux, Stephen, Edoardo Masset, Rachel Sabates-Wheeler, Michael Samson, Althea-Maria Rivas, and Dolf te Lintelo. 2017. "The Targeting Effectiveness of Social Transfers." *Journal of Development Effectiveness* 9 (2): 162–211. https://doi .org/10.1080/19439342.2017.1305981.

Drake, Lesly, Meena Fernandes, Elisabetta Aurino, Josephine Kiamba, Boitshepo Giyose, Carmen Burbano, Harold Alderman, Lu Mai, Arlene Mitchell, and Aulo Gelli. 2017. "School Feeding Programs in Middle Childhood and Adolescence." In *Disease Control Priorities: Child and Adolescent Health and Development*, third edition, edited by Donald A. P. Bundy, Nilanthi de Silva, Susan Horton, Dean T. Jamison, and George C. Patton, volume 8, 147–64. Washington, DC: World Bank. https://doi .org/10.1596/978-1-4648-0423-6_ch12.

Drèze, Jean, Prankur Gupta, Reetika Khera, and Isabel Pimenta. 2018. "Casting the Net: India's Public Distribution System after the Food Security Act (May 13, 2018)." https://ssrn.com/abstract=3177691 or https://www.doi.org/10.2139 /ssrn.3177691.

Durán-Valverde, Fabio, José F. Pacheco-Jiménez, Taneem Muzaffar, and Hazel Elizondo-Barboza. 2020. "Financing Gaps in Social Protection: Global Estimates and Strategies for Developing Countries in Light of the COVID-19 Crisis and Beyond." Working Paper No. 14, International Labour Organization, Geneva. https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---soc_sec/documents /publication/wcms_758705.pdf.

ECLAC (Economic Commission for Latin America and the Caribbean). 2002. *Meeting the Millennium Poverty Reduction Targets in Latin America and the Caribbean*. Santiago, Chile: ECLAC. https://repositorio.cepal.org/bitstream/handle/11362/2349/1 /S029628_en.pdf.

Filgueira, Fernando, and Ernesto Espindola. 2015. "Toward a System of Basic Cash Transfers for Children and Older Persons: An Estimation of Efforts, Impacts and Possibilities in Latin America." Políticas Sociales Series No. 216 (LC/L.3934), Economic Commission for Latin America and the Caribbean, Santiago, Chile. https://repositorio.cepal.org/handle/11362/37820.

Filmer, Deon, and Norbert Schady. 2008. "Getting Girls into School: Evidence from a Scholarship Program in Cambodia." *Economic Development and Cultural Change* 56: 581–617. http://hdl.handle.net/10986/5737.

Fuchs, Alan, Mariano Sosa, and Matthew Wai-Poi. 2021. "Progressive Domestic Resource Mobilization for a COVID-19 Recovery." Policy Brief, World Bank, Washington, DC.

Gaspar, Vitor, David Amaglobeli, Mercedes Garcia Escribano, Delphine Prady, and Mauricio Soto. 2019. "Fiscal Policy and Development: Human, Social, and Physical Investment for the SDGs." IMF Staff Discussion Note 19/03, International Monetary Fund, Washington, DC. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2019/01/18/Fiscal-Policy-and-Development-Human-Social-and-Physical-Investments-for-the-SDGs-46444.

Gelbach, Jonath, and Lant Pritchett. 1997. "More for the Poor Is Less for the Poor: The Politics of Targeting." Policy Research Working Paper 1799, World Bank, Washington, DC. https://imagebank2.worldbank.org/search/438833.

Gelli, Aulo, Elisabetta Aurino, Gloria Folson, Daniel Arhinful, Clement Adamba, Isaac Osei-Akoto, Edoardo Masset, Kristie Watkins, Meena Fernandes, Lesley Drake, and Harold Alderman. 2019. "A School Meals Program Implemented at Scale in Ghana Increases Height-for-Age during Midchildhood in Girls and in Children from Poor Households: A Cluster Randomized Trial." *Journal of Nutrition* 149 (8): 1434–42. https://doi.org/10.1093/jn/nxz079.

Gentilini, Ugo, Mohamed Almenfi, Ian Orton, and Pamela Dale. 2020. "Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures." Living Paper (various issues), World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/33635.

Gentilini, Ugo, Margaret Grosh, Jamele Rigolini, and Ruslan Yemtsov, eds. 2019. *Exploring Universal Basic Income: A Guide to Navigate Concepts, Evidence, and Practices*. Washington, DC: World Bank. http://hdl.handle.net/10986/32677.

Gentilini, Ugo, Margaret Grosh, and Michal Rutkowski. 2019. "Understanding Coverage: What Does Universal Social Protection Really Mean?" Development Matters. Organisation for Economic Co-operation and Development, Paris. https://oecd-development-matters.org/2019/01/28/understanding-coverage-what-does-universal-social-protection-really-mean/.

Giuliano, Paola, and Antonio Spilimbergo. 2014. "Growing up in a Recession." *Review of Economic Studies* 81 (2): 787–817. https://doi.org/10.1093/restud/rdt040.

Graham, Carol. 2002. "Public Attitudes Matter: A Conceptual Frame for Accounting for Political Economy in Safety Nets and Social Assistance Policies." Social Protection Discussion Paper 0233, World Bank, Washington, DC. https://imagebank2.worldbank.org/search/2163649.

Haggard, Stephan, Robert R. Kaufman, and James D. Long. 2013. "Income, Occupation, and Preferences for Redistribution in the Developing World."

Studies in Comparative International Development 48: 113–40. https://www .proquest.com/docview/1357134225?accountid=27683.

Harms, Phillip, and Stefan Zink. 2003. "Limits to Redistribution in a Democracy: A Survey." *European Journal of Political Economy* 19: 651–68. https://www .sciencedirect.com/science/article/pii/S0176268003000211.

HelpAge. 2018. "Social Pensions Data Base." HelpAge, London. www.pension -watch.net.

HelpAge International. 2011. "Financing Social Pensions in Low- and Middle- Income Countries." Pension Watch, Briefings on Social Protection in Older Age, Briefing 4, HelpAge International, London. http://www.pension-watch .net/resources/pension-watch-briefing-series/.

Hickey, Sam, Tom Lavers, Miguel Niño-Zarazúa, and Jeremy Seekings. 2018. "The Negotiated Politics of Social Protection in Sub-Saharan Africa." Working Paper 2018/34, United Nations University World Institute for Development Economics Research, Helsinki, Finland. https://econpapers.repec.org/paper /unuwpaper/wp-2018-34.htm.

Holmemo, Camilla, Pablo Acosta, Tina George, Robert J. Palacios, Juul Pinxten, Shonali Sen, and Sailesh Tiwari. 2020. *Investing in People: Social Protection for Indonesia's 2045 Vision*. Jakarta, Indonesia: World Bank. http://hdl.handle .net/10986/33767.

Hoy, Chris, and Andy Sumner. 2003. "Gasoline, Guns and Giveaways: Is There New Capacity for Redistribution to End Three Quarters of Global Poverty?" Working Paper 433, Center for Global Development, Washington, DC. https://www .cgdev.org/sites/default/files/gasoline-guns-and-giveaways-end-three-quarters -global-poverty-0.pdf.

ILO (International Labour Organization). 2012. *Social Security for All: Building Social Protection Floors and Comprehensive Social Security Systems: The Strategy of the International Labour Organization*. Geneva: ILO. https://www.ilo.org/dyn/normlex /en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID:3065524.

ILO (International Labour Organization). 2018. "Financing Social Protection for the Future of Work: Fiscal Aspects and Policy Options." Prepared by Staff of The ILO for the G20 Framework Working Group. ILO, Geneva.

ILO (International Labour Organization). 2021a "Building Social Protection Systems: International Standards and Human Rights Instruments." ILO, Geneva.

ILO (International Labour Organization). 2021b. *World Social Protection Report 2020– 22: Social Protection at the Crossroads—In Pursuit of a Better Future*. Geneva: ILO. https://www.ilo.org/global/publications/books/WCMS_817572/lang--en /index.htm.

ILO–UNICEF (International Labour Organization–United Nations Children's Fund). 2019. "Towards Universal Social Protection for Children: Achieving SDG 1.3." ILO–UNICEF Joint Report on Social Protection for Children. ILO, Geneva. https://www.unicef.org/reports/towards-universal-social-protection -children-achieving-sdg-13.

IMF (International Monetary Fund). 2017. *Fiscal Monitor: Tackling Inequality*. Washington, DC: IMF. https://www.imf.org/en/Publications/FM/Issues/2017/10/05 /fiscal-monitor-october-2017.

IMF (International Monetary Fund). 2019. "Brazil Article IV Report." IMF, Washington, DC.

IMF (International Monetary Fund). 2021. *Fiscal Monitor Update*, January. Washington, DC: IMF. https://www.imf.org/en/Publications/FM/Issues /2021/01/20/fiscal-monitor-update-january-2021.

Inchauste, Gabriela, and Nora Lustig. 2017. "The Distributional Impact of Taxes and Transfers." World Bank, Washington, DC. http://hdl.handle.net/10986/27980.

ISSA (International Social Security Association). 2018. *Social Security Programmes throughout the World: Europe 2017.* Geneva: ISSA.

Iversen, Torben, and David Soskice. 2006. "Source Electoral Institutions and the Politics of Coalitions: Why Some Democracies Redistribute More Than Others." *American Political Science Review* 100 (2): 165–81. https://www.cambridge.org /core/journals/american-political-science-review/article/electoral-institutions -and-the-politics-of-coalitions-why-some-democracies-redistribute-more -than-others/F30F0B252475D90BFF365C2410FA23DB.

Jacques, Olivier, and Alain Noel. 2018. "The Case for Welfare State Universalism, or the Lasting Relevance of the Paradox of Redistribution." *Journal of European Social Policy* 28 (1): 70–85.

Jorgensen, Steen Lau, and Paul B. Siegel. 2019. "Social Protection in an Era of Increasing Uncertainty and Disruption: Social Risk Management 2.0." Social Protection and Jobs Discussion Paper 1930, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/31812.

Karabarbounis, Loukas. 2011 "One Dollar, One Vote." *Economic Journal* 121 (553): 621–51. https://www.jstor.org/stable/41236996?seq=1#metadata_info_tab_contents.

Klemm, Alexander D., and Paolo Mauro. 2021. "Pandemic and Progressivity." IMF Working Paper 21/24, International Monetary Fund, Washington, DC. https:// doi.org/10.5089/9781513568621.001.

Maluccio, John A., and Rafael Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social." Research Report 141, International Food Policy Research Institute, Washington, DC. https://ideas.repec.org/p/fpr/resrep/141.html.

McCord, Anna, and Rachel Slater. 2009. "Overview of Public Works Programmes in Sub-Saharan Africa." Overseas Development Institute, London.

Meltzer, Alan H., and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy* 89 (5): 914–27. https://doi .org/10.1086/261013.

Milanovic, Branko. 2000. "The Median-Voter Hypothesis, Income Inequality, and Income Redistribution: An Empirical Test with the Required Data." *European Journal of Political Economy* 16: 367–410. https://www.sciencedirect.com/science /article/pii/S0176268000000148.

Nowack, Daniel, and Sophia Schoderer. 2020. "The Role of Values for Social Cohesion: Theoretical Explication and Empirical Exploration." Discussion Paper 6/2020, German Development Institute, Bonn, Germany. https://www .die-gdi.de/uploads/media/DP_6.2020.pdf.

OECD (Organisation for Economic Co-operation and Development). 2019. *Can Social Protection Be an Engine for Inclusive Growth?* Development Centre Studies. Paris: OECD Publishing. https://doi.org/10.1787/9d95b5d0-en.

Olinto, Pedro, Gabriel Lara Ibarra, and Jaime Saavedra-Chanduvi. 2014. "Accelerating Poverty Reduction in a Less Poor World: The Roles of Growth and Inequality." Policy Research Working Paper 6855, World Bank, Washington, DC. http://hdl .handle.net/10986/18347.

Olivera, Javier. 2015. "Preferences for Redistribution in Europe." *IZA Journal of European Labor Studies* 4 (Article 4772). https://doi.org/10.1186/s40174 -015-0037-y.

Oliviera, Victor, Laura Tiehen, Mark Prell, and David Smallwood. 2018. "Evolution and Implementation of the Supplemental Nutrition Program in the United States." In *The 1.5 Billion People Question*, edited by Harold Alderman, Ugo Gentilini, and Ruslan Yemtsov, 209–64. Washington, DC: World Bank.

Ortiz, Isabel, Matthew Cummins, and Kalaivani Karunanethy. 2017. "Fiscal Space for Social Protection Options to Expand Social Investments in 187 Countries." Extension of Social Security Series Working Paper No. 48, International Labour Office, Geneva, and UNICEF and UN Women, New York.

Ostry, Jonathan D., Andrew Berg, and Charalambos G. Tsangarides. 2014. "Redistribution, Inequality, and Growth." IMF Discussion Note 14/02, International Monetary Fund, Washington, DC. https://www.imf.org/external /pubs/ft/sdn/2014/sdn1402.pdf.

Packard, Truman G., Ugo Gentilini, Margaret Ellen Grosh, Philip B. O'Keefe, Robert J. Palacios, David A. Robalino, and Indhira Vanessa Santos. 2019. *Protecting All: Risk Sharing for a Diverse and Diversifying World of Work*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/997741568048792164 /Protecting-All-Risk-Sharing-for-a-Diverse-and-Diversifying-World-of-Work.

Ralston, Laura, Colin Andrews, and Allan Hsiao. 2017. "The Impacts of Safety Nets in Africa: What Are We Learning?" Policy Research Working Paper 8255, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986 /28916.

Rodriguez, Laura, and Matt Wai-Poi. 2020. "Fiscal Policy, Poverty and Inequality in Jordan: The Role of Taxes and Public Spending." Middle East and North Africa Poverty Practice, World Bank, Washington, DC.

Saez, Emmanuel, and Gabriel Zucman. 2019. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. W. W. Norton & Company. ISBN:9781324 002734.

Sehnbruch, Kirsten, and Rafael Carranza. 2015. "The Chilean System of Unemployment Insurance Savings Accounts." Serie de Documentos de Trabajo 401, Departamento de Economia, Universidad de Chile, Santiago. https:// www.researchgate.net/publication/278678166_The_Chilean_System _of_Unemployment_Insurance_Savings_Accounts.

Sepúlveda, Magdalena. 2016. "Ensuring Inclusion in Social Assistance Programs: The Role of the Human Rights Principle of Equality and Non-Discrimination." Prepared for the Project Getting It Right: Learning How to Translate Human Rights Principles into Social Protection Policy. World Bank, Washington, DC.

Sepúlveda, Magdalena. 2018. "What Does Human Rights Law Have to Do with Social Protection Systems? Operationalizing the Human Rights Principles of Equality and Transparency." UN Women Expert Group Meeting, Sixty-Third Session of the Commission on the Status of Women, UN Women, New York.

https://socialprotection.org/discover/publications/what-does-human-rights-law-have-do-social-protection-systems-operationalizing.

Sepúlveda, Magdalena, and Carly Nyst. 2012. "The Human Rights Approach to Social Protection." Ministry of Foreign Affairs of Finland, Helsinki, Finland. https://www.ohchr.org/Documents/Issues/EPoverty/HumanRightsApproach ToSocialProtection.pdf.

Soares, Sergei, Letícia Bartholo, and Rafael Guerreiro Osorio. 2019. "Uma Proposta para a Unificação dos Benefícios Sociais de Crianças, Jovens e Adultos Pobres e Vulneráveis." Discussion Paper 2505, Instituto de Pesquisa Economica Aplicada, Rio de Janeiro, Brazil.

Sparrow, Robert. 2004. "Protecting Education for the Poor in Times of Crisis: An Evaluation of a Scholarship Program in Indonesia." SMERU Working Paper, SMERU Research Institute, Jakarta, Indonesia. https://www.smeru.or.id/sites /default/files/publication/jpseducation.pdf.

Subbarao, Kalanidhi, Carlo del Ninno, Colin Andrews, and Claudia P. Rodriguez Alas. 2013. *Public Works as a Safety Net: Design, Evidence, and Implementation.* Directions in Development: Human Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/11882.

Taylor-Gooby, Peter. 2005. *Attitudes to Social Justice*. London: Institute for Public Policy Research. https://www.ippr.org/files/images/media/files/publication /2011/05/attitudes_social_justice_1350.pdf.

UNICEF (United Nations Children's Fund). 2019. "Programme Guidance: Strengthening Shock Responsive Social Protection Systems." UNICEF, New York. https://www.unicef.org/documents/programme-guidance-strengthening -shock-responsive-social-protection-systems.

UNICEF–ODI (United Nations Children's Fund–Overseas Development Institute). 2020. "Universal Child Benefits: Policy Issues and Options." UNICEF–ODI, New York, https://www.unicef.org/media/72916/file/UCB-ODI-UNICEF -Report-2020.pdf.

UNRISD (United Nations Research Institute for Social Development). 2013. "A Framework for Implementing a Human Rights-Based Approach to Social Protection." UNRISD, Geneva. https://socialprotection-humanrights.org /framework/.

van Oorschot, Wim. 2006. "Making the Difference in Social Europe: Deservingness Perceptions among Citizens of European Welfare States." *Journal of European Social Policy* 16 (1): 23–42.

van Oorschot, Wim, and Femke Roosma. 2015. "The Social Legitimacy of Differently Targeted Benefits." ImPRovE Working Paper No. 15/11, Herman Deleeck Centre for Social Policy, University of Antwerp, Belgium.

World Bank. 2006. *World Development Report 2006: Equity and Development.* Washington, DC: World Bank. http://hdl.handle.net/10986/5988.

World Bank. 2012. *Resilience, Equity, and Opportunity: The World Bank's Social Protection and Labor Strategy 2012–2022*. Washington, DC: World Bank Group. http:// documents.worldbank.org/curated/en/443791468157506768/Resilience -equity-and-opportunity-the-World-Banks-social-protection-and-labor -strategy-2012-2022.

World Bank. 2016. "Financing Social Protection." Background paper for the East Asia and Pacific Social Protection Regional Report/Memo. World Bank, Washington, DC.

World Bank. 2018a. *Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle*. Washington, DC: World Bank. https://openknowledge.worldbank.org /bitstream/handle/10986/30418/9781464813306.pdf.

World Bank. 2018b. *The State of Social Safety Nets 2018*. Washington, DC: World Bank. http://hdl.handle.net/10986/29115.

World Bank. 2019. *World Development Report 2019: The Changing Nature of Work*. Washington, DC: World Bank. http://hdl.handle.net/10986/30435.

World Bank. 2020a. "Brazil: Income Support for the Poor Affected by COVID-19 Project." World Bank, Washington, DC.

World Bank. 2020b. *Poverty and Shared Prosperity 2020: Reversals of Fortune*. Washington, DC: World Bank. http://hdl.handle.net/10986/34496.

World Bank. 2021a. *Global Economic Prospects*, January. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/34710.

World Bank. 2021b. "Investing in Human Capital for a Resilient Recovery: The Role of Public Finance." Human Capital Project, World Bank, Washington, DC. http://hdl.handle.net/10986/35840.

World Bank and DFAT (Australian Department of Foreign Affairs and Trade). Forthcoming. *Diverse Paths: The Dynamic Evolution of Social Protection in Asia and the Pacific*. Washington, DC, and Sydney, Australia: World Bank and DFAT.

Yemtsov, Ruslan, Aleksandra Posarac, Maria Nagernyak, and Irina Albegova. 2019. "Towards More Effective Social Assistance in Russia: An Update of the System Performance Considering New National Target of Halving Poverty by 2024." World Bank, Washington, DC. https://openknowledge.worldbank.org /handle/10986/3225.

2

Unpacking the Empirics of Targeting in Low- and Middle-Income Countries

Margaret Grosh, Claudia P. Rodriguez Alas, Usama Zafar, Matthew Wai-Poi, Emil Tesliuc, and Phillippe Leite

To help illuminate the choices around whether to differentiate eligibility or benefits across the welfare distribution, it is important to understand not just the theory, but also the empirics of the trade-offs involved. There are many parts to the whole empirical story.

This chapter provides a broad overview of the coverage, incidence, and simple[1] estimates of the impacts on poverty of a wide range of social assistance programming in developing countries, using the Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE)[2] global data set. It casts a wide net to look at emerging and developing countries with recent household survey data and all kinds of social assistance programs, irrespective of the intention to delimit the benefits or eligibility thresholds or the methods used to do so. The analysis provides benchmarking for the outcomes observed, which can be useful comparators for country-specific discussions and setting expectations for the feasibility of different scenarios. For selected observations, the data set supplements the broad picture that is observable from the survey of the programs. Thus, the chapter uses the data set to contrast program choices and outcomes, foreshadowing the deeper discussion on such choices in the subsequent chapters. The focus of the ASPIRE database is monetary welfare, whether income or consumption. Chapter 3 discusses wider concepts of welfare.

This chapter also provides a brief synopsis of the costs of targeting. To address disincentive costs, the chapter summarizes the conceptual issues and a few of the seminal works from the wider impact evaluation literature. The summary is brief because the literature is extensive and needs no addition. On administrative costs, the chapter outlines the issues in measurement and provides illustrative numbers from a range of programs. The evidence on transaction costs, stigma, and social costs that is available in the literature is much scarcer, but it is enough to signal the importance of such costs and efforts to reduce them. Altogether, the evidence on administrative costs, transaction costs, and stigma motivates chapter 4 on delivery systems and how they can/should be improved. Essay 9, in chapter 1, covers the political costs of targeting.

Other chapters bring in other parts of the empirical story. Chapter 5 illustrates how simulations can help in comparing geographic, demographic, and household-specific targeting methods. It also catalogs the nascent literature on experiments with comparative treatment arms among household-specific methods, mostly comparing community-based methods and proxy means testing. Such experiments are important because community-based targeting is difficult to simulate credibly. Chapter 6 provides the evidence base—especially from simulations and deep process evaluations—that illuminates how each targeting method can be designed to best advantage.

This chapter focuses on the most proximate indicators for targeting in the Besley and Kanbur (1990) framework—coverage, benefit levels, incidence, and simple estimates of changes in poverty. Policy makers or social policy analysts may differ in the weights they give to these outcomes vis-à-vis others, but all programs have distributional outcomes. For countries with pledges to reduce poverty and inequality and provide universal social protection within this decade, the chapter discusses a relevant although not comprehensive set of indicators.[3]

Measurement and Interpretation

Measurement across the Welfare Spectrum

The traditional framing of the targeting problem as dichotomous is too simplistic. In a dichotomous framing, the value of a transfer is the same for anyone below the poverty line/eligibility threshold, and it is zero for anyone above the threshold. It has long been recognized that in most people's minds and in most social welfare functions, a more continuous valuation makes sense. A transfer to someone at the very bottom of the welfare distribution, say the 5th centile, seems more important than a transfer to someone in the 19th centile. The value of a transfer to someone in the 19th centile does not seem to be much different from the value of a transfer to someone in the

21st centile, even if the 20th centile is eligibility threshold. Even a transfer to someone in the 30th centile would be valued more than a transfer to someone in the 60th centile. Dichotomous measures exaggerate the notion of targeting errors by placing as much value on an error in distinguishing between someone on the 19th or 21st centiles as an error in distinguishing someone in the 5th or 60th centiles. In a continuous welfare function, the first kind of error, just around a threshold, is much more frequent but not nearly as important as the second kind of error among people with very different levels of welfare, which occurs less frequently.

The continuous welfare function has important consequences for measurement—giving preference to more continuous measures of coverage and incidence or summative measures, such as impacts on poverty or over the whole distribution rather than the dichotomous formulation of errors of inclusion and exclusion. This theme is taken up in greater detail in chapter 7. Although the limitations of dichotomous framing have long been understood, the use of the dichotomous measures of errors of inclusion and exclusion is still surprisingly prevalent.

A difficulty in evaluating the targeting problem is in weighting the errors of inclusion versus the errors of exclusion. In poverty economics, this problem is often solved by the choice of the welfare function. For example, a common choice is to use a Foster-Greer-Thorbecke poverty headcount (FGT0) measure for a dichotomous welfare function and an FGT poverty gap (FGT1) or squared poverty gap (FGT2) for a continuous version. These measures can be applied to simulations of different transfer levels, budgets, selection rules, and the implied targeting errors. Such exercises help in exploring trade-offs (Acosta, Leite, and Rigolini 2011; Brown, Ravallion, and van de Walle 2016; Knox-Vydmanov 2011, embedded in the ADePT software[4]). Other ways to weight one kind of error against another are used and described in chapters 6, 7, and 8.

The form of the welfare function determines how much more value is put on the welfare of the very poorest versus that of the just poor, the not quite poor, or those who are better off. Most such continuous welfare functions imply that there is some tolerance for errors of exclusion. If a few of the intended population are excluded but enough extra value can be given to the poor(-er) who are included, it will reduce overall poverty or inequality and thus be judged better for society as a whole.

For policy makers, the available budget can influence the relative importance of one type of error versus the other in a more practical way. Consider two thought experiments:

- *With a firmly binding budget constraint.* This scenario starts with a baseline of whatever poverty levels have gone before the initiation of a possible new poverty-targeted program, which will at least initially be budget rationed to include fewer than all the poor in the country. In such a

scenario, a poor person who is excluded by a targeting error ends up with the same welfare[5] as those who are unserved because the program is small. A needy person served by the program is better off. And any non-poor person who is included (error of inclusion) takes up budget that could have helped a poor person in the target group. In this scenario of budget rationing, reducing errors of inclusion is a means to reduce errors of exclusion, and errors of exclusion are inevitable given the rationing. This scenario fits well the situations in poor countries that are just start-ing to build their social protection systems. Beegle, Coudouel, and Monsalve (2018) provide several African examples using the latest data available at the time of the study. In Ethiopia, the poverty rate (measured by the international standard of $1.90 purchasing power parity/day) was 34 percent, but social assistance covered only 8 percent of the population. In Kenya, the poverty rate was also 34 percent, but social assistance cov-ered only 6 percent of the population. In Tanzania, the poverty rate was 47 percent and social assistance covered 13 percent of the population. These three countries have significant, new flagship programs. In coun-tries such as Sierra Leone or Madagascar, the disjuncture was much larger—in Sierra Leone, there were more than 10 times as many poor people as those served by social assistance; in Madagascar, it was more than 20 times. In such cases of budget rationing, including nonpoor peo-ple in the programs will crowd out the poor. To cover all the poor, increas-ing budgets is vital, but reducing errors of inclusion will help as well.

- *With a less binding budget constraint.* An alternative scenario starts with a budget that is sufficient at least to serve all those who are poor plus any nonpoor who are in the program by design or due to errors in eligibility assessment. In this case, as before, reducing errors of exclusion is vital to ending poverty and realizing the principle of nondiscrimination as artic-ulated in the human rights frameworks. However, the ability to do so is not rationed by the budget but by potential deficiencies in the delivery system or targeting mechanism. Reducing an error of inclusion may save budget, but with a budget that is already sufficient to serve all the poor, it will not map directly to reducing errors of exclusion.

The increasing use of human rights rather than economic perspectives and the prevalence of the goal of universal social protection seem to be building a consensus that errors of exclusion must be given greater weight. However, the concern about errors of exclusion is hardly new, having been prominently flagged years ago in the United Nations Children's Fund's (UNICEF) work on the social costs of adjustment and calls for adjustment with a human face (see, for example, Cornia, Jolly, and Stewart 1987). Economic welfare functions that place heavier weight on the welfare of the poor than the less poor or nonpoor are consonant with the consensus that errors of exclusion are important.

Pros and Cons of Household Survey–Based, Cross-Country Comparisons

Looking at a wide range of countries and programs can help in understanding the degree to which there are common findings or marked variability. Understanding the degree to which inferences can be made helps to establish realistic expectations.

Household survey–based analysis of targeted social programs is highly sensitive to the method used, which puts a premium on being able to use primary data that can be handled with comparable methods. Results are sensitive to how welfare aggregates are constructed—whether households are ranked by welfare using the pre- or posttransfer welfare, to poverty lines, to eligibility thresholds, and so forth. The World Bank invested in the ADePT SP software[6] and ASPIRE[7] to improve comparability across some of these dimensions, and the analysis in this chapter relies on these strengths.

However, general purpose household surveys such as those captured in ASPIRE may miss a portion of social assistance programming. The surveys are usually designed for multiple purposes and often rooted in providing weights for the consumer price index. Such surveys may not have samples among the poor that are large enough or questionnaires that are well-tuned to pick up participation in social protection programs, especially small ones. The problem may be most acute for low-income countries where survey data tend to be scarcer and social protection programs have only recently emerged at scale. Until programs are well cemented in national policy and large enough to observe systematically in survey samples, it is unsurprising that questionnaire designers would not alter their traditional questionnaires, especially since a large body of survey design practice shows that measurements are sensitive to changes in instruments. Chapter 7 provides more on these issues, including some examples.

For international benchmarking, it is usual to report on metrics that are useful for cross-country comparisons but that may be different from those used for specific countries or programs. There are two especially common aspects of this.

- *The threshold.* ASPIRE analysis often discusses the poorest quintile, as this chapter does as well. But, as discussed in chapter 1, eligibility thresholds for different programs can be set for smaller or larger shares of the population, and poverty rates will differ as well. In high-income countries, the poverty line or eligibility threshold for guaranteed minimum income programs may be set lower than the bottom quintile. For example, Chile designed its Chile Solidario program for the bottom 5 percent of the population because that was the share that was chronically poor by the

country's standards. In a low-income country (LIC), a program with a low budget may aim to cover only the poorest, say, 10 percent of the population at the start, even if this is a lower threshold than the absolute poverty line. Even if such programs perfectly meet their program-specific goals, the programs would show under-coverage in a cross-country comparison using the lowest quintile as the target group. To help counterbalance this problem, the next section presents findings for the whole distribution rather than focusing solely on the bottom quintile as the only threshold.

• *Ancillary eligibility criteria.* Simple cross-country, cross-program benchmarking does not account for all the other characteristics or criteria that influence eligibility for a given program. For example, many social assistance programs are for children (child grants and school feeding programs) or families with children (conditional cash transfers and many other unconditional ones). However, a nontrivial share of households do not have children in the eligible age range (a much lower share in Africa than in Eastern Europe due to differences in demography and the frequency of multigenerational households) and conversely for programs dedicated to the elderly. To help counterbalance this problem, the next section presents some findings for the whole of social assistance programming, rather than program by program.

To allow cross-country comparability, the comparisons use a common welfare aggregate, which differs from the welfare aggregate used by some of the programs being compared. The income or consumption used in the analysis could differ from the country-specific or program-specific definition (for example, not accounting for specific types of income). The welfare aggregate is harmonized for differences in the size of the household in a per capita indicator, while some programs may use an adult equivalence scale to determine the operational welfare aggregate used for eligibility and/or define multiple assistance units within the household. Because of these factors, the harmonized welfare aggregate does not coincide with the operational welfare aggregate that some individual programs use, causing a downward bias in the "true" benefit incidence of those programs.

More precise assessments of misclassification in eligibility assessments in a specific program are specialized and conducted with a variety of methods. For means- or asset-tested programs, the assessments may rely on re-interviews of households, more extensive cross-checks with other databases, and triangulation between what is reported in application files and what is seen in representative national surveys (see box 6.7 in chapter 6 for an illustration). For proxy means tests or geographic targeting, ex ante simulations are often used to show how well the algorithms select from the larger pool represented in household surveys (see discussions in chapters 5, 6, and 7). For community-based methods, the results may be compared

with household survey–type information (or special samples of the poor), although the comparisons must acknowledge that communities' definitions of needs and those used in the surveys to assess welfare may be different, a theme taken up in chapter 6.

Despite all these caveats on the interpretation of the results, looking at household survey data is too useful to forgo, so that is done in the next section. Other papers use a range of country surveys to assess targeting performance across multiple contexts (for example, Kidd and Athias 2020). The next section extends this type of cross-country analysis using the ASPIRE database, which has the advantages of using harmonized survey data to enhance comparability and large country and program coverage. In addition, the section looks at a broader range of performance indicators than is commonly used. Kidd and Athias (2020) focus on exclusion errors, which are an important but incomplete view of targeting outcomes (see chapter 7). This chapter looks at coverage across the whole distribution, incidence across the whole distribution, the benefit level and adequacy of benefits, and changes in the poverty gap.[8]

Recent ASPIRE Survey–Based Evidence of Targeting Outcomes

This section presents ASPIRE's main distributional performance indicators, including coverage, incidence, and impact on poverty and inequality. From the larger set of 432 surveys covering 125 countries over the past two decades, the chapter focuses on results from 2014 onward, essentially the most recent five years of data on the platform (when the data were drawn for this compilation in June 2021, only one survey for 2019 was available in ASPIRE).[9] This yields a sample of 70 countries and their most recent surveys: 24 in Sub-Saharan Africa, 18 in Latin American and the Caribbean, 16 in Europe and Central Asia, 7 in East Asia and Pacific, 4 in South Asia, and 1 in the Middle East and North Africa. Annex 2A provides a complete list of the surveys. By income group, the sample has 7 high-income countries (all World Bank borrower clients, not traditional donor high-income countries), 28 upper-middle-income countries, 26 lower-middle-income countries, and 9 low-income countries.[10] Use of the 2014 cutoff achieves a reasonably broad coverage of countries. The average year of survey was 2016 in East Asia and Pacific, Europe and Central Asia, the Middle East and North Africa, and Sub-Saharan Africa and 2017 in Latin American and the Caribbean and South Asia. As social assistance programming was a very fast-moving field even before COVID-19, older data in countries with nascent or reforming programs are not representative of social protection today, but history can still be a useful teacher.

ASPIRE estimates the distributional performance of the social protection program categories based on household survey data and information on the sizes of the programs (number of beneficiaries and spending) and design parameters (including the targeting methods used for eligibility determination) from its administrative database. The two databases are not yet linked at the program level, which limits the ability to use the information on targeting methods (from the administrative database) with the information on distributional performance (from the household survey database). Chapter 5 presents information on the prevalence of different targeting methods, alone or mixed, from the administrative database. As of June 2021, ASPIRE captures 2,623 individual social assistance transfer programs in its administrative database. The household survey database includes only social assistance program categories, aggregating the individual social assistance programs recorded in household surveys. As of June 2021, there were 857 individual social assistance programs captured in the most recent surveys of the 125 countries covered by ASPIRE's household survey database. Although these represent only 30 percent of the programs captured in the administrative database, they tend to be the largest programs focused on households.[11]

In ASPIRE, indicators are presented for eight categories of social assistance programs: unconditional cash transfers, conditional cash transfers, social pensions, food and in-kind transfers, school feeding, labor intensive public works, fee waivers and targeted subsidies, and other social assistance. Where there are multiple programs observed in the questionnaire within a single category, the results for these programs are combined within the category. For example, a country may have both a "universal" age-based child allowance and a guaranteed minimum income program identified in the survey questionnaire. Since they are both unconditional cash transfers, they are reported on a single line despite having very different approaches to eligibility and presumably different program-specific results. Grouping the data facilitates the processing of hundreds of surveys in uniform and automatable ways. The results help in understanding world social assistance programming in aggregate, but the grouping is less helpful for understanding the results of specific programs or the potential of different targeting methods.[12]

The collation by category can be sensitive, depending on the programs that have been aggregated. For Ukraine, for example, there are six different unconditional cash transfer (UCT) programs observed in the questionnaire for the Household Living Conditions Survey. Three of these programs differentiate eligibility by welfare level (help for low-income families, help for single mothers, and assistance for children under guardianship or care) and three only by demographic characteristics/age of the children in the family (childbirth help, childcare benefit for children younger than three years,

and other child benefits). The childbirth help program is far larger than the others, as shown in figure 2.1, panel a. Nearly 22 percent of the people in Ukraine live in households that receive a UCT, 15 percent live in families that receive childbirth help, and the other programs cover 0.5 to 4 percent of the population. Although all the programs show a progressive trend,

Figure 2.1 **Coverage and Beneficiary Incidence of Unconditional Cash Transfer Programs in Ukraine and Mongolia as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare**

a. Coverage in Ukraine

- - All unconditional cash transfers
- - Childbirth help
- - Help for low-income families
- - Other child support
- - Help for single mothers with children
- - Child care allowance at the age of 3 years
- - Assistance for children under guardianship or care
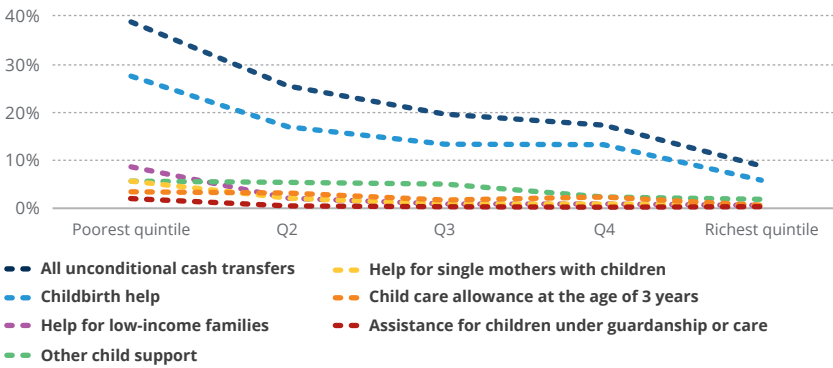
b. Beneficiary incidence in Ukraine

Share of beneficiaries
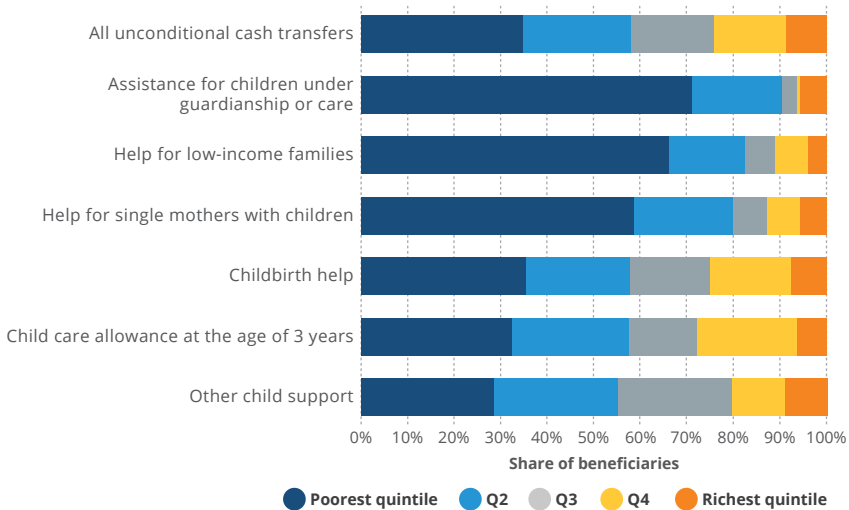
Poorest quintile   Q2   Q3   Q4   Richest quintile

Figure 2.1 *(continued)*

c. Coverage in Mongolia

- All unconditional cash transfers
- Child allowances
- Mother benefits
- Funeral
- Human development
- Other government transfers
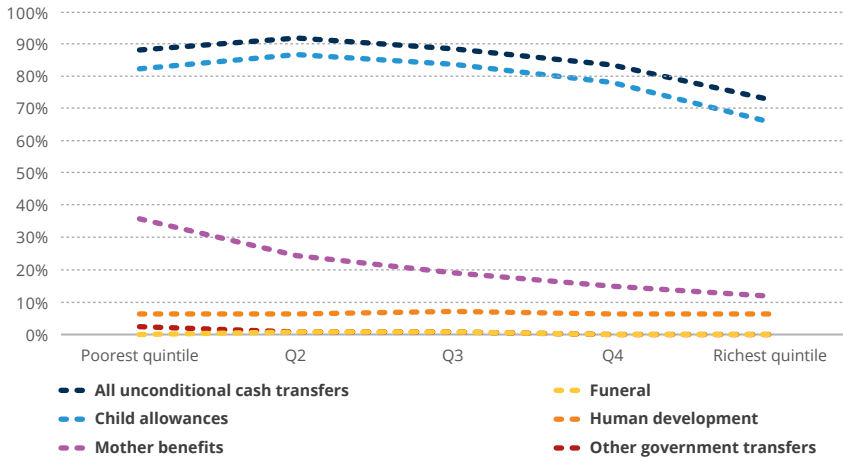
d. Beneficiary incidence in Mongolia
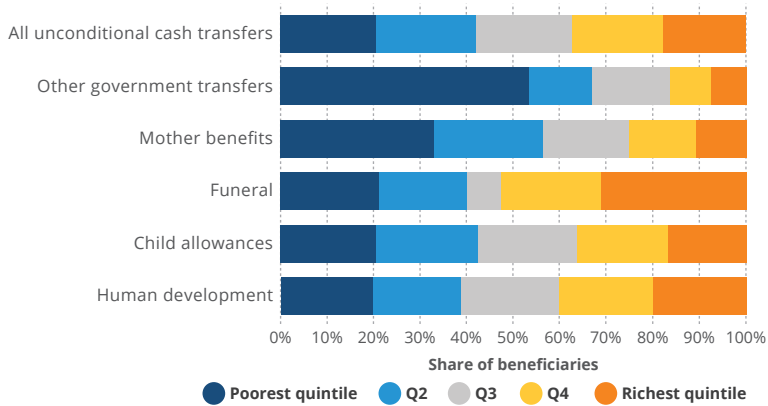
- Poorest quintile
- Q2
- Q3
- Q4
- Richest quintile

Source: ASPIRE database, www.worldbank.org/aspire.

Note: Figure based on unconditional cash transfers (UCT) programs captured in Ukraine's 2016 Household Living Condition Survey and Mongolia's 2016 Household Socio-Economic Survey. UCTs include any of the following: poverty alleviation and emergency programs; guaranteed minimum income programs; and universal or poverty-targeted child and family allowances. They do not include social pensions or targeted subsidies in cash. Coverage is: (Number of individuals in a given group [for example, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

figure 2.1, panel b, illustrates that the income-targeted programs have a greater share of beneficiaries in the poorest quintile. However, the bar for all unconditional cash transfers tracks closely the coverage of the largest program. This effect is even more marked for Mongolia, where the relative weight of the (then) universal child allowance program (Child Money Program) in the category is even higher. Overall, 85 percent the population lives in households that receive a UCT, and 79 percent live in households that receive the Child Money Program. As figure 2.1, panels c and d, shows, the overall picture for unconditional cash transfers tracks closely that of the Child Money Program.

These data are not sufficient to answer questions about which targeting methods work best. To do that well would require comparing a large number of programs that cover the range of targeting methods, have excellent documentation on the design and implementation of each program, and have primary data on outcomes. Ideally, this would be done over a large range of contexts. However, ASPIRE does not provide data that are rich enough for such an exercise: the ASPIRE data group programs according to category rather than providing program-specific data; targeting methods cannot be easily assigned to the programs, in part because a third of the programs use multiple methods and the importance of each is impossible to untangle; and the institutional variables are missing and of limited richness in an important share of the cases. Even if such data were available, it would be difficult to draw simple and clean conclusions. The larger is the number of programs considered, the less is the institutional richness that can be brought to the story to sort out differences in intent (how important is poverty reduction or redistribution among a program's objectives?), setting (especially the degree of poverty and inequality), design (where the eligibility threshold is set, the indicators of need used, and the structure of benefits or services), or execution (quality of implementation and challenges along the delivery chain, as explained in chapter 4). This chapter discusses these features for a few examples. Chapters 3 to 6 focus on them in greater detail.

The data can help in understanding the outcomes achieved by the world's social protection programming, which in turn helps to set realistic expectations and understand why certain themes are important in the targeting literature. The incomplete coverage explains the emphasis in the social protection sector on universal social protection writ large and the concern about errors of exclusion in targeted programs more specifically. The variation in the incidence results found across programs suggests that policy choices matter—that is, at least in some circumstances, it is possible to focus an important share of resources at the bottom end of the distribution. The variation also shows that even the results with the sharpest targeting do not achieve "perfect" targeting, not all the benefits reach the

poorest, and not all the poorest are covered. This explains the emphasis in the literature and dialogue on understanding the sources of both sorts of errors and considering their implications, the policy weight given to changing each outcome and the options for improving the delivery chain, the details of the targeting method selected, and/or the selection of targeting method.

The next sections use this framework and information from the ASPIRE data set to look at recent data on program coverage, incidence, the interplay of coverage and incidence, and the relationship of these proximate factors with the overall impact on poverty. The sections use the ASPIRE data set to examine these proximate factors and the overall impact on poverty across different program groups and countries.

Coverage

Coverage by social assistance is 36 percent for the world as a whole and 54 percent for the poorest quintile. Moreover, coverage increases with country income but with high dispersion (figures 2.2 and 2.3). The box and whisker plots in figure 2.2 show the means value within each box, the interquartile range as the colored box, and the range as the ends of the whiskers. In low-income countries, on average, only 17 percent of people in the poorest quintile are covered by social assistance, whereas the figure is 77 percent for high-income countries.[13] Coverage of the full population by social assistance is lower than coverage of the poorest quintile, at 36 and 54 percent, respectively, across all countries, indicating that by and large, countries have chosen and implemented progressive programs. The difference in coverage of the whole population and the poorest quintile is fairly small in low-income countries, at 13 and 17 percent, respectively. The difference is much more substantial for other income groups. Upper-middle-income countries, for example, cover 41 percent of their whole population with social assistance and 64 percent of their poorest quintile. The range of coverage overall is quite broad, from virtually none to virtually all the poorest quintile being covered, but there are truncations of the range at the extremes of the country income groups. None of the low-income countries manages to cover more than about half the poorest quintile; none of the World Bank client high-income countries covers less than half of the poorest quintile. The box and whisker plot is useful for capturing quotable headline facts, but a scatterplot may be more appropriate to see the patterns (figure 2.3).

Country income level is not destiny in social assistance coverage—clearly, policy choices play a role. This is confirmed looking at a scatter plot of the data on social assistance coverage of all types of programs for the

Figure 2.2   Coverage of the Total Population and Poorest Quintile Receiving Social Assistance as Captured in the ASPIRE Household Surveys, by Income Group
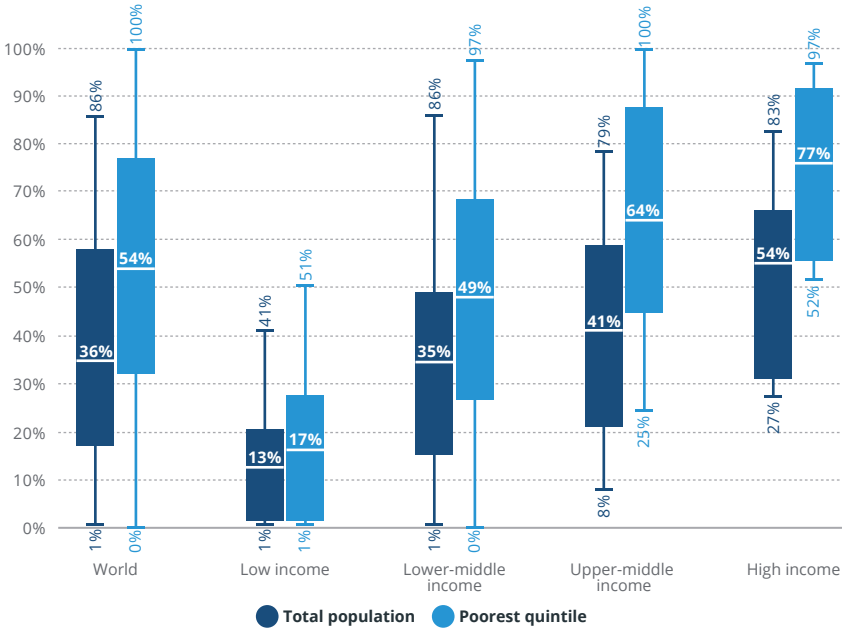
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: World (n = 70), low income countries (n = 9), lower-middle-income countries (n = 26), upper-middle-income countries (n = 28), high income countries (n = 7). Aggregated indicators are calculated using simple averages of country-level social assistance coverage rates across regions. Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption).

poorest quintile by gross domestic product (GDP) per capita at the country level, which again shows wide variability. The range of country income that the World Bank labels lower-middle-income countries is relatively narrow, from about $1,000 to $4,000 per capita, but it shows the full range of social assistance coverage of the poorest quintile from zero to nearly complete. The range of income covered by the upper-middle-income country label is much greater (from about $4,000 to $12,500 per capita), but it still shows significant variation in coverage of the poorest quintile. Coverage of the poorest quintile is between 25 percent (Montenegro [MNE]) up to almost

Figure 2.3 Scatterplot of Social Assistance Coverage for the Poorest Quintile, as Captured in the ASPIRE Household Surveys, and GDP per Capita (US$)
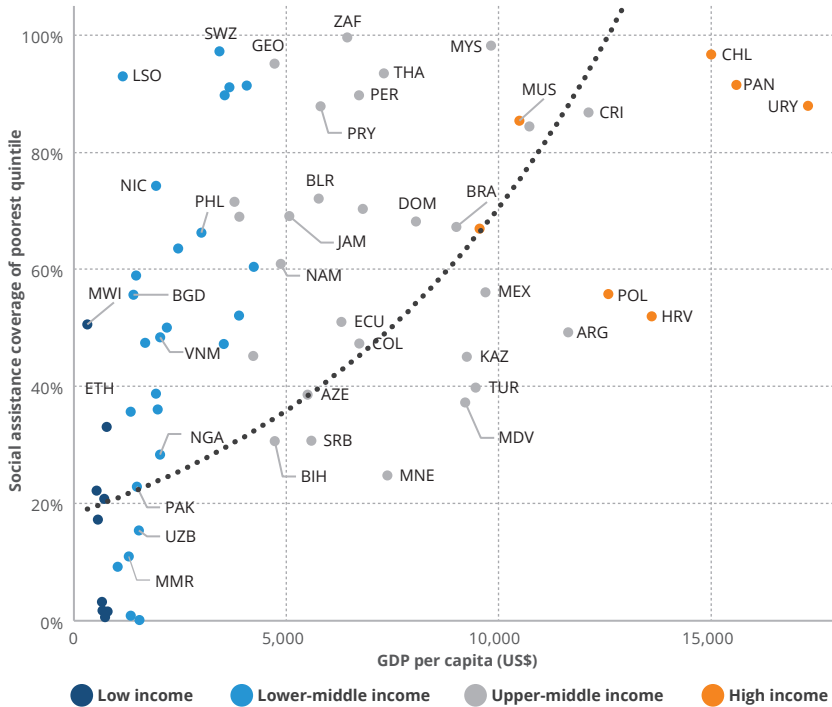
Note: The number of countries per income group is as follows: Total (n = 70), low income countries (n = 9), lower-middle-income countries (n = 26), upper-middle-income countries (n = 28), high income countries (n = 7). Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). GDP per capita (US$) is obtained from World Development Indicators (WDI) for the same year as each country's survey. This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption). Annex 2A lists the country codes; GDP = gross domestic product.

100 percent (South Africa [ZAF]). The high-income countries included in the sample still show variability in coverage rates, but the dispersion is smaller—from 52 percent (Croatia [HRV]) to 97 percent (Chile [CHL]) of the poorest quintile. The lowest coverage of the poor among high-income countries is higher than the highest coverage rate reported by low-income countries. Despite the variability, the overall trend of the function is upward, meaning that coverage rates of the poor tend to increase with country income.

About 43 percent of the countries (27 of 63) in the sample cover so few people with social assistance that even with perfect targeting there would be significant errors of exclusion. All the low-income countries have total coverage rates below the share of the poor in the population, and all the high-income countries cover a greater share. Figure 2.4 plots the poverty headcount from PovCalNet for each country using the standard World Bank suite of international poverty lines ($1.90/day for low-income countries, $3.20/day for lower-middle-income countries, and $5.50/day for upper-middle-income countries and high-income countries) against coverage of the population. Countries that are below the 45-degree line have total coverage of social assistance that is too low to cover all the poor. In Niger (NER), for example, 45.4 percent of people are poor by the $1.90/day line, but only 20.1 percent are covered by social assistance. In Kenya (KEN), 66 percent of people are poor by the $3.20/day line, but only 26 percent are covered by social assistance. In Ecuador (ECU), 24 percent of the population is poor by the $5.50/day poverty line, but 20 percent are covered by some type of social assistance program. Conversely, 60 percent of the countries provide social assistance to more than the share of the population deemed poor.

Higher total coverage is positively but imperfectly correlated with higher coverage of the poor. As shown in figure 2.5, almost all the observations fall above the 45-degree line, that is, they cover a higher share of the poorest quintile than of the overall population.[14] This is consistent with all the prior graphs, but figure 2.5 shows how strong the pattern is and how dispersion in the coverage of the poorest quintile increases with coverage of the whole population. Program types with low coverage overall, at less than around 5 percent of the population, show a range of coverage of the poorest quintile, from about 5 percent (proportionate) to covering about 25 percent of the poorest quintile (sharply targeted). Similarly, for country program categories covering about 15 percent of the population, the share of the poorest quintile covered ranges from about 20 to about 45 percent. Where coverage reaches about 30 percent of the population overall, the coverage of the poorest quintile spans the range from about 35 to about 80 percent of the poorest quintile. So, while larger programs that are closer to universal naturally find it easier to cover the poorest, many smaller programs also provide significant coverage to the poorest.

Incidence

Aggregating across countries, the results for all sorts of social assistance programs show progressive incidence (figure 2.6). The largest share of beneficiaries, ranging from about 30 to 50 percent of total beneficiaries depending on the program type, are in the poorest quintile. Each successive quintile

Figure 2.4 Scatterplot of Social Assistance Coverage for the Total Population and the Poverty Headcount Ratio, as Captured in the ASPIRE Household Surveys
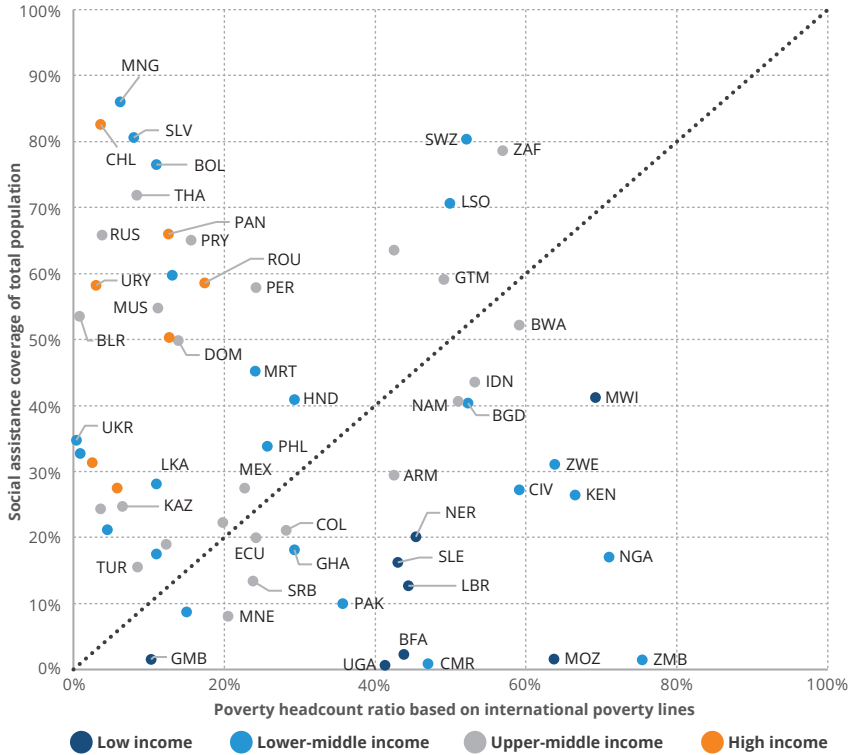
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per income group is as follows: Total (n = 63), low income countries (n = 8), lower-middle-income countries (n = 24), upper-middle-income countries (n = 24), high income countries (n = 7). Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Poverty rates are obtained from World Bank PovCalNet for the same year as each country's survey and based on international poverty lines for the relevant country income group; US$1.90 per person per day (2011 PPP) for low income countries, US$3.20 for lower-middle-income countries and US$5.50 for upper-middle-income countries and high income countries. Poverty rates were not available for Azerbaijan, Bosnia and Herzegovina, Ethiopia, Jamaica, Malaysia, Tanzania, and Uzbekistan. This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. Annex 2A lists the country codes. PPP = purchasing power parity.

Figure 2.5   Coverage of Social Assistance (Poorest Quintile versus Total Population), as Captured in the ASPIRE Household Surveys, by Program Type
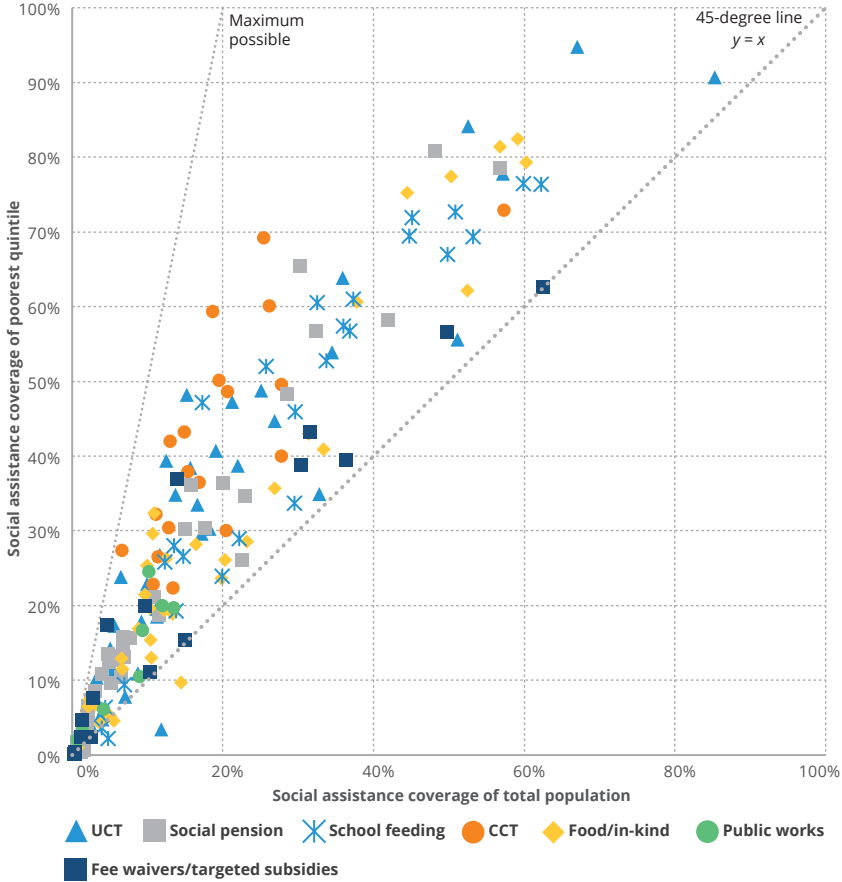
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of programs is as follows: unconditional cash transfers (n = 49), conditional cash transfers (n = 20), social pensions (n = 38), food and in-kind transfers (n = 36), school feeding (n = 26), public works (n = 8), fee waivers and targeted subsidies (n = 17). Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/ (Number of individuals in the group). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption). UCT = unconditional cash transfer; CCT = conditional cash transfer.

Figure 2.6 **Global Distribution of Beneficiaries, as Captured in the ASPIRE Household Surveys, by Type of Social Assistance Instrument and Quintile of Pretransfer Welfare**
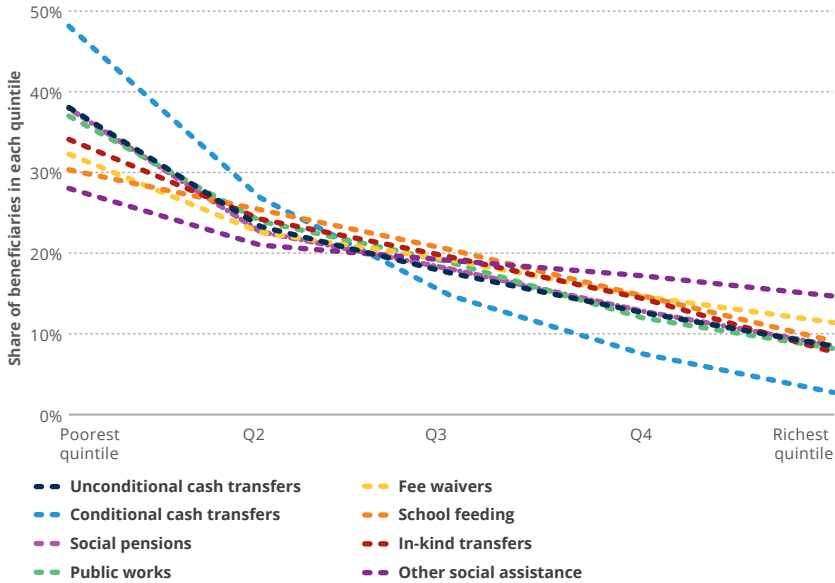
Legend:
- Unconditional cash transfers
- Conditional cash transfers
- Social pensions
- Public works
- Fee waivers
- School feeding
- In-kind transfers
- Other social assistance

Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of programs is as follows: unconditional cash transfers (n = 49), conditional cash transfers (n = 20), social pensions (n = 38), public works (n = 8), fee waivers and targeted subsidies (n = 17), school feeding (n = 26), in-kind transfers (n = 36), other social assistance (n = 47). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Aggregated indicators are calculated using simple averages of program instrument beneficiaries' incidence rates across countries. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

receives less, down to 10 percent or less for the richest quintile for each of the program types. Such downward-sloping incidence curves are typical of program- and country-specific analysis in the wider literature. There are two reasons why conditional cash transfers have the steepest incidence curves. First, in most countries with conditional cash transfers represented in the household survey, there is a single program in the category and the program is clearly poverty-targeted. Second, many countries have multiple individual programs in the UCT category with varied designs, as in the example of Ukraine. School feeding is the category with the least progressive incidence, and these programs are usually targeted geographically and by age group commensurate with the schools covered but very rarely by methods that distinguish between the welfare levels of students within a given school.

Focusing on country-specific results reveals a great deal of variation in the share of benefits accruing to the poorest quintile, irrespective of the type of program. For example, figure 2.7 shows the share of beneficiaries in each quintile for each country for which unconditional cash transfers are observed in the household survey data. From 6 to 73 percent of the beneficiaries[15] of unconditional cash transfers are in the poorest quintile of the population. Annex 2B provides analogous graphs for other types of programs. They too show large variation, although the range and number of observations are largest for unconditional cash transfers. This may reflect the wide variety of programming that falls under this categorization, as well as the varied degree to which the surveys capture the programs.

Mapping of Coverage against Incidence

Univariate descriptions are interesting for benchmarking, but a richer understanding comes from considering country and program specifics and multiple factors at once. Figure 2.8 maps the coverage of the poorest quintile against the share of beneficiaries in the poorest quintile. The figure shows the same strong dispersion on each axis that the univariate figures did but there is no strong pattern or relationship between the two, nor across program types. The "targeting nirvana" of covering all the poorest and concentrating most of the benefits on the poorest would be found in the upper right quadrant of this graph, an area rather empty. Of course, countries might intend to serve less than 20 percent of the population, so a country-nuanced reading might call something less than 100 percent coverage of the poorest quintile successful. Thus, it is useful to explore the quadrants using cases where there happens to be only one program or a strongly dominant program within a "program category," to bring in some backstory about the specific program.

Uruguay's (URY) UCT is near the origin of the graph, far from the ideal targeting goal on both dimensions. In figure 2.7, Uruguay has the least progressive incidence in the UCT category. In Uruguay, there is only one program observed in the UCT category, the Prima por Hogar Constituido (Transfer for Constituted Household).[16] Only 6 percent of the program's beneficiaries are in the poorest quintile. The program is provided only to public servants who are married or have dependents and whose monthly gross salary is less than two times the national minimum wage. The program might be called "affluence tested" because it has a high threshold, but by virtue of hinging on public formal sector employment, it excludes most of the poor, covering only 3.5 percent of the poorest quintile and accruing to a member of the household for 12 percent of the population overall. Mongolia's (MNG) UCT is at the far right of the graph. In this case, the UCT category is dominated by the Child Money Program, which is well-known

Figure 2.7   Distribution of Unconditional Cash Transfer Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare
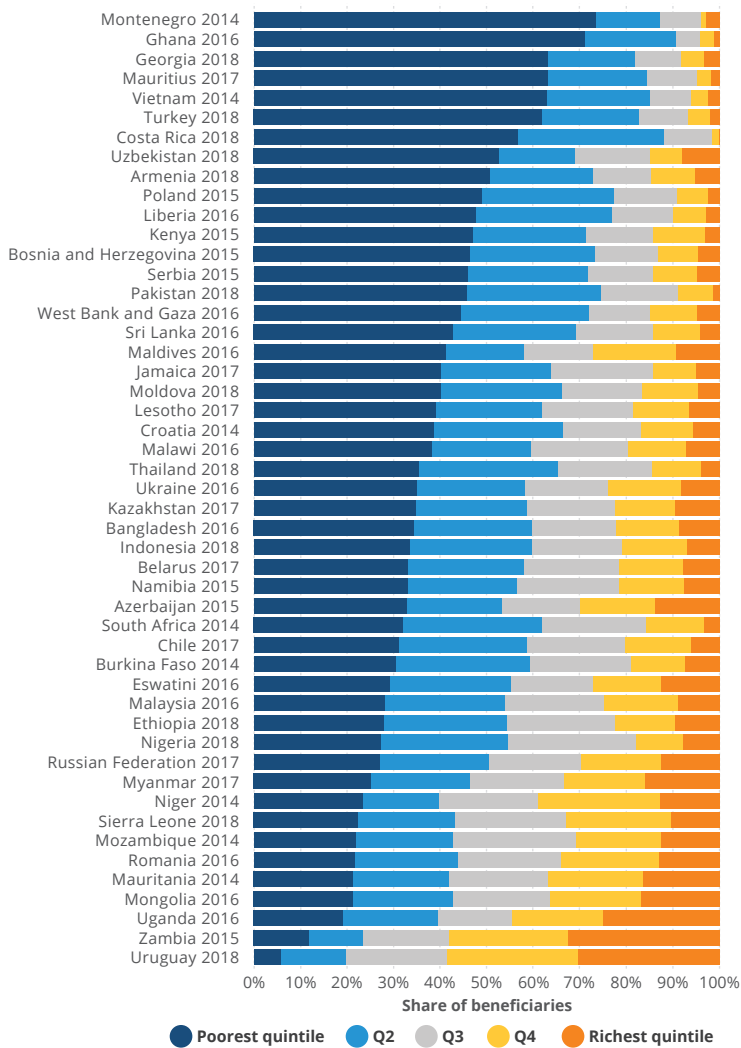
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region with unconditional cash transfers is as follows: Total (n = 49), Sub-Saharan Africa (n = 18), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 4), East Asia and Pacific (n = 6), South Asia (n = 4), Middle East and North Africa (n = 1). Unconditional cash transfers include any of the following: poverty alleviation and emergency programs; guaranteed minimum income programs; and universal or poverty-targeted child and family allowances. They do not include social pensions or targeted subsidies in cash. Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

in the social assistance literature. Mongolia's UCT shows very high coverage of the poorest quintile and nearly perfectly proportional incidence (as seen in figure 2.7). This result is congruent with the (then) universal design of the Child Money Program, which was financed with revenues from natural resource extraction (ILO 2016; UNICEF 2020).[17] Its flat incidence is the flip side of its high coverage: 91 percent of the poorest quintile and 85 percent of the population living in recipient households, in a country with a poverty rate of 6.2 percent.[18] Montenegro (MNE) is at the other end of the spectrum of incidence of unconditional cash transfers. Its unconditional cash transfer category is dominated by its guaranteed minimum income program (Gotcheva et al. 2013; Republic of Montenegro 2013).[19] The UCT category delivers 73 percent of benefits to the poorest quintile but covers only a quarter of the households in the quintile and 6.5 percent of the population, although by the international poverty line of $5.50/day, Montenegro's poverty headcount is 21 percent. In Ghana (GHA), the Livelihood Empowerment against Poverty program is the only UCT observed. It is progressive (71 percent incidence of the poor) and targeted through a combination of geographic, demographic, and proxy means testing. However, it was a small program at the time of the survey, covering only 5 percent of the poorest quintile and 1.5 percent of the population, although the poverty headcount was 30 percent (Republic of Ghana 2018). (All poverty headcounts are calculated using international rather than local poverty lines.)

Overall Effect on Poverty

To reduce poverty, social programs need to cover a significant number of poor households. Such coverage can be driven by concentrating benefits on the poor, increasing the size of the programs, or both. Figure 2.9 shows how the program coverage of the poorest quintile (represented by the size of the bubble) depends on the share of beneficiaries found in the poorest quintile (y-axis) and the size of the program (x-axis).[20] The x-axis is truncated to exclude very large programs, to focus on the more common range of program sizes, from very small to those covering around 40 percent of the population. The analysis will include very large programs next. The first set of programs in the left-hand grouping have medium to highly progressive incidence but are very small in total size, covering less than 10 percent of the total population. Despite their highly progressive incidence, the small size means that their coverage of the poorest quintile cannot be high; if all the beneficiaries of a program representing 5 percent of the population were in the poorest quintile, there would be only 25 percent coverage of the first quintile. The middle and right-hand groupings show programs with higher levels of coverage.

Figure 2.8   Coverage and Incidence of Social Assistance Beneficiaries in the Poorest Quintile, as Captured in the ASPIRE Household Surveys, by Country and Type of Program
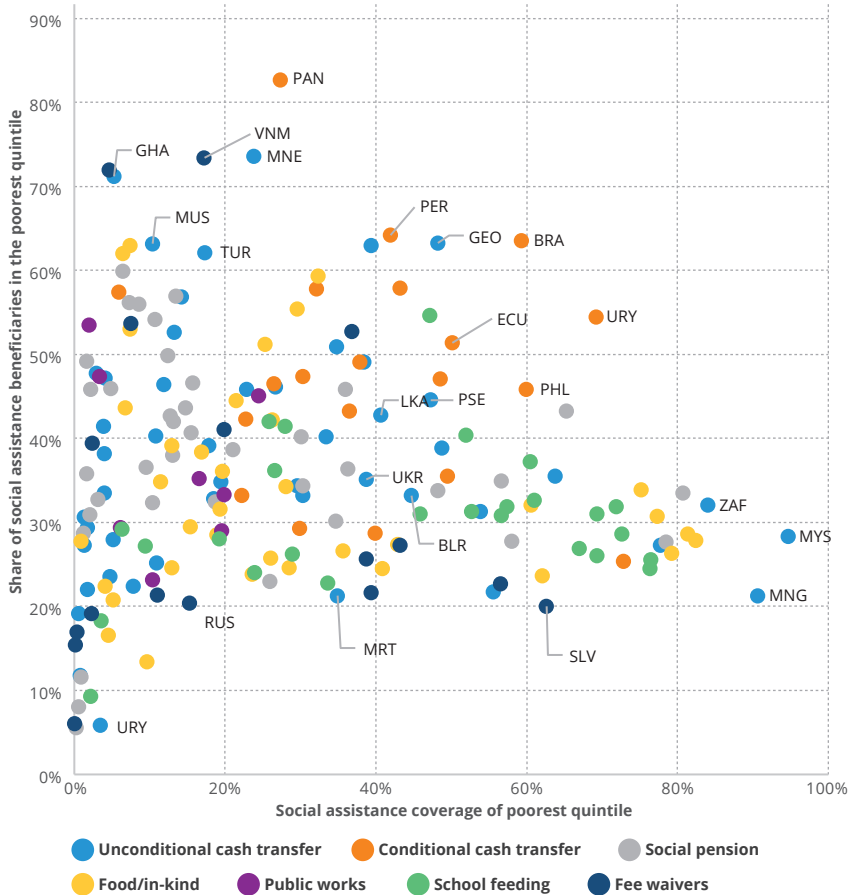
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: World (n = 70), Sub-Saharan Africa (n = 24), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 18), South Asia (n = 4), Middle East and North Africa (n = 1). The number of programs is as follows: unconditional cash transfers (n = 49), conditional cash transfers (n = 20), social pensions (n = 38), food/in-kind (n = 36), public works (n = 8), school feeding (n = 26), fee waivers (n = 17). Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption). Annex 2A lists the country codes.

Figure 2.9   Incidence and Coverage of Cash Program Beneficiaries, as Captured in the ASPIRE Household Surveys, Excluding Very Large Programs
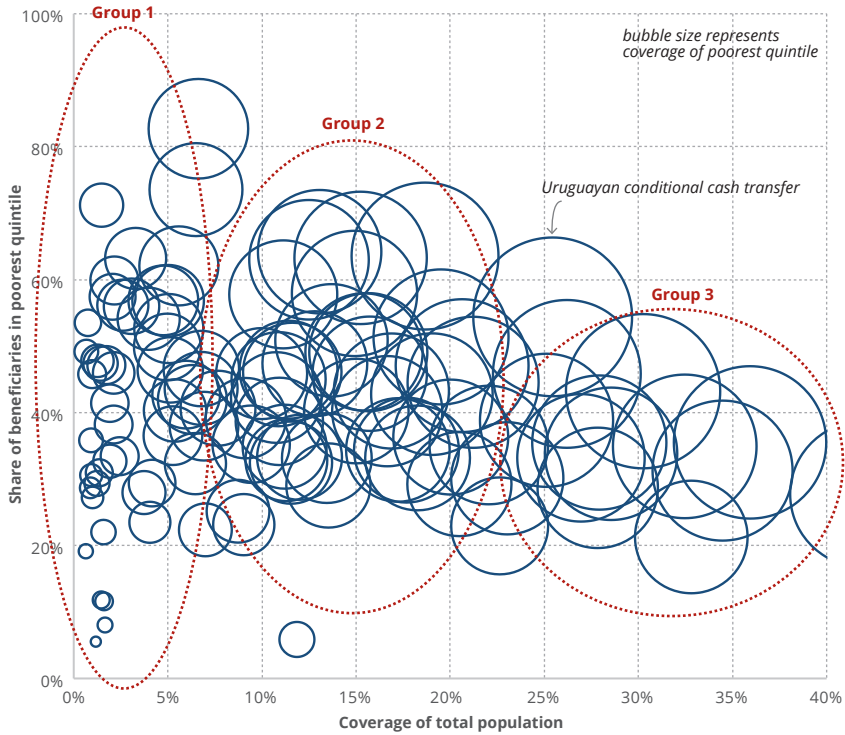
Groups 1–3 from left to right: **Group 1. Highly progressive incidence but small programs in total size; Group 2. Moderate progressive incidence in medium-size programs; Group 3. Less progressive incidence but large program sizes.**

Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region with cash programs is as follows: World (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 49), conditional cash transfers (n = 20), social pensions (n = 38), public works (n = 8). Coverage is: (Number of individuals in a given group [for example, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption).

The middle group achieves coverage of the poorest quintile with medium-size programs and moderately progressive incidence. The right-hand group achieves coverage of the poorest quintile through larger program size but less progressive incidence. When a larger program can achieve a highly progressive incidence, coverage of the poorest quintile is particularly high, as seen in the case indicated by the arrow, which is the Uruguayan conditional cash transfer program Asignaciones Familiares,[21] which concentrates 54 percent of beneficiaries in the poorest quintile, covering 25 percent of all households and 69 percent of those in the poorest quintile.

Once a program exceeds a certain size, the share of benefits that can accrue to the poorest decreases. In figure 2.9, there is high variation of incidence for the small programs (properly speaking, these are program/categories, but here they are referred to as "programs"). The variation is smaller for the medium-size programs and smaller again for the larger programs. In a sense, this is mechanical; when a program exceeds the population of the poor, the share of poor beneficiaries must be less than 100 percent, and as programs grow larger, the share of the beneficiaries who are poor must continue to fall even when 100 percent of them are covered. That downward-sloping trend is somewhat evident in figure 2.9 and becomes starker in figure 2.10, which is on the same axes but covers the full range so that the handful of very large programs covering more than 40 percent of the population are included. For the largest program/category, the UCT in Mongolia, 85 percent of the population lives in households covered by the program, and 91 percent of people in the poorest quintile live in households covered by the program. Together, figures 2.9 and 2.10 show that (1) very small programs can never achieve high coverage of the poor because they are too small; (2) greater coverage of the poor can be achieved through larger programs, more progressive incidence, or both; and (3) for particularly large programs, the coverage of the poor will generally be relatively high, but the share of benefits going to the poor will be relatively low.

While increasing the share of benefits accruing to the poor, the size of a program, or both can increase coverage of the poor, increasing the concentration of benefits among the poor is the most cost-effective way to reduce poverty. To understand the relative roles of incidence and program size, figure 2.11 overlays on a single grid the share of beneficiaries in the poorest quintile and the total program coverage of the population, both plotted against the poverty gap reduction per $1 spent (a standard output from ADePT). Thus, each program observation appears on the graph twice, in different colors. For example, in the upper left and upper right of the

Figure 2.10   Incidence and Coverage of Cash Program Beneficiaries, as Captured in the ASPIRE Household Surveys, Including Very Large Programs

Source: ASPIRE database, www.worldbank.org/aspire.

Notes: The number of countries per region with cash programs is as follows: world (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 49), conditional cash transfers (n = 20), social pensions (n = 38), public works (n = 8). Coverage is calculated as (number of individuals in a given group [for example, total population or poorest quintile] who live in a household in which at least one member receives the transfer)/(number of individuals in the group). Beneficiary incidence is calculated as (number of direct and indirect beneficiaries [people who live in a household in which at least one member receives the transfer] in a given quintile)/(total number of direct and indirect beneficiaries). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption).

Figure 2.11   Efficiency of Cash Programs, as Captured in the ASPIRE Household Surveys

Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region with cash programs is as follows: World (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 43), conditional cash transfers (n = 15), social pensions (n = 35), public works (n = 6). Monetary information was not available for 16 programs to generate poverty gap indicators. For this reason, the sample of cash programs used for this figure is smaller than the one used for figures estimating coverage and incidence of benefits. Coverage is: (Number of individuals in a given group [that is, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption). PAN = Panamanian conditional cash transfer.

figure, at the poverty gap value of 0.8, two dots are labeled PAN. These represent the Panamanian conditional cash transfer program category, which is dominated by the Red de Oportunidades program, a smallish but fairly traditional Latin American conditional cash transfer program.[22] The 0.8 means that the poverty gap is reduced by 80 cents for each dollar spent on the program. The dark blue dot (coordinate 82.7, 0.8) shows that about 83 percent of the beneficiaries are in the poorest quintile, which is the most progressive observation in the sample. However, the total program coverage is modest. The light blue dot (coordinate 6.6, 0.8) shows that the programs covers just under 7 percent of the total population (and therefore just 27 percent of the poorest quintile).

There is a much stronger relationship between the incidence of benefits and reduction in the poverty gap per dollar spent than between the overall coverage of the population and reduction in the poverty gap per dollar spent. The blue line representing incidence of benefits for the poorest quintile increases relatively steeply, indicating that as a greater share of program benefits goes to the poorest quintile, the degree of poverty gap reduced for each dollar spent increases. Specifically, an additional 10 percent share to the poorest quintile means the poverty gap falls by an extra $0.10 per $1 spent. Conversely, the light blue line representing total program coverage of the population rises only modestly until programs reach around 30–40 percent population coverage, and then it declines. Although there is high variation among smaller programs, ranging from best to worst performers in terms of efficiency of poverty reduction, the largest programs reduce the poverty gap the least per dollar spent.

Poverty reduction depends not only on poor households receiving social benefits, but also on those benefits being large enough to be meaningful. High coverage of poor households will not reduce poverty very much if the level of benefits that they receive is very small. Figure 2.12 shows the coverage of the poorest quintile (x-axis) and adequacy of benefits (benefits as a percentage of posttransfer welfare, y-axis), with the reduction in the poverty gap shown by the size of the bubble. Both axes are initially truncated to exclude programs with very high adequacy or very high coverage of the poor, to focus on the range where practice is concentrated. These restrictions are relaxed next. Again, there are three groups. The adequacy of benefits for the first group (on the left) varies significantly, from under 10 percent of consumption (or income in the case of Latin America and the Caribbean and the Russian Federation) to over 50 percent. Nonetheless, the degree of poverty gap reduction is relatively small because only 15 percent or less of the poorest quintile is covered. The second group (in the middle at the bottom) has higher levels of coverage, but the adequacy

Figure 2.12 Coverage of the Poorest Quintile, Program Adequacy, and Poverty Reduction of Cash Programs, as Captured in the ASPIRE Household Surveys, Excluding High-Coverage Programs
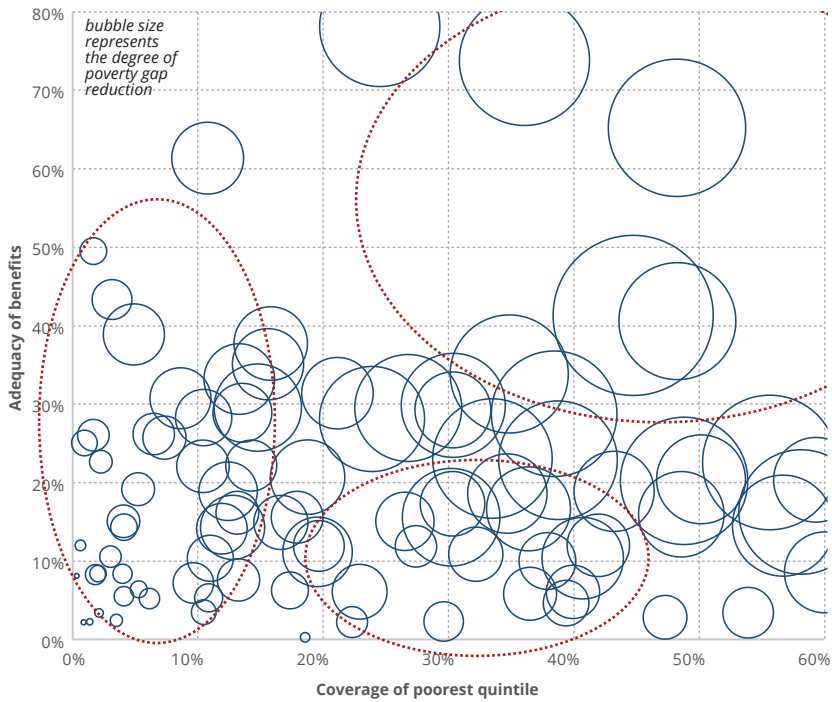
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region with cash programs is as follows: World (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 43), conditional cash transfers (n = 15), social pensions (n = 35), public works (n = 6). Monetary information was not available for 16 cash programs to generate adequacy and poverty gap indicators. For this reason, the sample of programs used for this figure is smaller than the one used for figures estimating coverage and incidence of benefits. Coverage is: (Number of individuals in a given group [for example, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Adequacy is the mean transfer amount received by a given group (for example, poorest quintile) as a share of the total welfare of the beneficiaries in that group. This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption).

levels are below 20 percent, meaning that reduction of the poverty gap is somewhat higher than for the first group, but less than that of the third group (in the middle at the top), which has medium coverage but high adequacy.

A combination of good coverage of the poor and good adequacy of benefits is the most effective way to reduce the poverty gap. Figure 2.13 includes high coverage and very high adequacy programs. As before, programs with low coverage and adequacy do little to reduce the poverty gap. Programs with very high coverage but low adequacy do a little better, but not as much as those with good coverage and good adequacy. Finally, some extreme outliers combine very high coverage with very high adequacy. They achieve the largest reductions in the poverty gap, but the very high benefits may imply a disincentive to work (unless there are countervailing design parameters), and the combination of very high benefits and coverage means that the programs will require large budgets. For example, the bubble flagged by the pointer in the figure is unconditional cash transfers in South Africa. That category is dominated by the country's Child Support Grant (see figure 2C.1, panel b, in annex 2C). The program is means tested with a relatively high threshold and a substantial benefit for families with children up to age 18, with a budget in 2014 of 1.0 percent of GDP. The other unconditional cash transfers account for a further 0.25 percent of GDP. The Child Support Grant has been evaluated to have substantial impacts on a wide range of indicators of use of services and outcomes for health, nutrition, and education for children along the age spectrum from birth to adolescence but not to have negative effects on adult labor (see CGD 2015; DSD, SASSA, and UNICEF 2012; Plagerson and Ulriksen 2015).

Although coverage of the poor and adequacy of benefits both drive poverty reduction, they also raise questions of cost and efficiency. The group of high program coverage of the poorest quintile in figure 2.13 achieves reasonable poverty reduction despite low adequacy. This is because there are so many people in the poorest quintile who live just below the poverty line that even a relatively small benefit level can bring many of them close to or even above the poverty line. Figure 2.10 shows that many of these programs with large coverage of the poorest quintile have very large population coverage. Thus, although these programs achieve good poverty reduction, in many cases, they do so at a relatively high fiscal cost due to their large population coverage, and inefficiently due to their low share of benefits going to the poorest quintile. Policy makers need to balance program coverage with program adequacy and budget implications to determine what will have the greatest impact on poverty.

Figure 2.13 **Coverage of the Poorest Quintile, Program Adequacy, and Poverty Reduction of Cash Programs, as Captured in the ASPIRE Household Surveys, Including High-Coverage Programs**

Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: World (n = 65), Sub-Saharan Africa (n = 20), East Asia and Pacific (n = 7), Europe and Central Asia (n = 16), Latin America and the Caribbean (n = 17), South Asia (n = 4), Middle East and North Africa (n = 1). The number of cash programs is as follows: unconditional cash transfers (n = 43), conditional cash transfers (n = 15), cocial pensions (n = 35), public works (n = 6). Monetary information was not available for 16 programs to generate adequacy and poverty gap indicators. For this reason, the sample of cash programs used for this figure is smaller than the one used for figures estimating coverage and incidence of benefits. Coverage is: (Number of individuals in a given group [for example, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Adequacy is the mean transfer amount received by a given group (for example, poorest quintile) as a share of the total welfare of the beneficiaries in that group. This figure underestimates total social assistance coverage because household surveys do not include all programs existing in each country. The poorest quintile is calculated using per capita pretransfer welfare (income or consumption).

Evidence Base for the Costs of Poverty Targeting

There has been a gradual accretion of evidence on the costs of poverty targeting. There is an increasingly firm body of knowledge on labor disincentives in the impact evaluation literature. There is much less in quantity and rigor on the other topics because the main data sources are process evaluations, which tend to be buried in the archives of program documents rather than the stuff of journal articles, but the available evidence is fairly consistent.

Labor Disincentives

Theory and intuition supports the notion that targeting that reduces benefits as a household's or individual's earnings rise could decrease work effort. In richer countries with programs that tend to use means testing, benefit differentiation, and sometimes offer significant levels of benefits, incentive issues are a noticeable feature of concern in the literature and policy debates. This is especially the case where families may be eligible for multiple programs, each with its own means test or sliding scale of benefits. Moffitt (2015) compiles evidence on the largest programs in the United States. The marginal tax rates across individual programs vary and also across income levels. For example, the Supplemental Nutrition Assistance Program (commonly known as food stamps) has a nominal 30 percent marginal tax rate, but it is effectively 24 percent because of earnings exclusion provisions. The Earned Income Tax Credit generates a marginal tax rate as high as −45 percent at the bottom of the scale, but it is 21 percent in the phaseout range. Cumulative marginal tax rates for families in the Supplemental Nutrition Assistance Program and facing federal and state income and payroll taxes, which implicitly include the Earned Income Tax Credit and child tax credits, show a range depending on family composition, earnings, number of workers, and so forth. For families with earnings below 50 percent of the poverty line, the marginal tax rate varies from −3 to 35 percent, with a median of 13 percent. For families with earnings between 150 and 200 percent of the poverty line, the marginal tax rates range from 22 to 51 percent, with a mean of 31 percent. The empirical evidence on impacts on work effort shows no significant effects overall but some for single-mother households.

So far in developing countries, few programs have a combination of features that would trigger a high level of concern about work effects and/or there are countervailing features at work in labor decisions. Many programs do not determine eligibility based only or principally on current earnings. Few adjust benefits at all as income rises, and those that do tend to do so with one or two steps or with earnings disregards. Few programs

reassess eligibility frequently. Many programs pay such low benefits that households that are capable of work effort have plenty of incentive to work to increase their incomes. Moreover, the evaluation evidence shows that transfers can release constraints to work. For example, by paying regularly, the programs may help households buy inputs they need for their farms or microenterprises and thus make work more productive, or they may make it easier for households to afford the resources needed for job search.

Indeed, a body of evidence built over the past decade around work incentives shows that the concern is overblown with respect to social assistance in developing countries. For example, in Bastagli et al.'s (2016) review of 74 studies on cash transfers, for just over half of the studies reporting on adult work, the cash transfer does not have a statistically significant impact. Among the studies reporting a significant effect among working-age adults, the majority find an increase in work participation and intensity. In the cases where a reduction in work participation or work intensity is reported, it reflects a reduction in participation among the elderly or those caring for dependents or is linked to reductions in casual work. Banerjee et al. (2017) analyze data from randomized controlled trials of cash transfer programs in six developing countries and find no systematic evidence that the programs discourage work. Baird, McKenzie, and Özler (2018) provide a narrative review of the extensive literature and find that prime-age adults show very little change in the amount they work or the amount they earn when receiving unconditional cash transfers, conditional cash transfers, or charitable grants. Transfers that enable people to find jobs in different places and to start new businesses have resulted in more labor and higher income for the recipients.

Although it is more nascent, the behavioral economics literature suggests various ways in which social assistance may improve work effort or its fruits. There is fairly conclusive evidence that financial concerns can reduce mental bandwidth and thereby cognitive capacity and executive control, with implications for risk taking and decision making, and that these affect the poorer more than the less poor (Schilbach, Schofield, and Mullainathan 2016). If the causality runs in the opposite direction as well, that a bit more money via a transfer can unlock bandwidth, then transfers may raise productivity. The evidence on these effects is still nascent. In an experimental study in India, Kaur et al. (2018) vary the timing of payments to workers doing piece-rate work. Those who received their pay early, and thus were less pressed about money, increased their hourly output and had fewer attentional errors. In a possibly proximate chain from transfers to psychosocial welfare to behavior, Attah et al. (2016) show that cash transfers in four African countries had positive impacts on psychosocial well-being, which led to further positive impacts on educational performance, participation in social life, and empowerment for decision making. In Pakistan,

Kosec and Mo (2017) study the effects of extremely heavy rainfall and widespread flooding during the 2010 monsoon. They conclude that social protection not only restored livelihoods and replaced damaged assets, but also had an enduring effect by easing mental burdens and raising aspirations for the future.

Administrative Costs

An often cited argument against differentiating eligibility or benefits by welfare is that household-specific eligibility assessments require prohibitively higher administrative costs than programs that differentiate eligibility by characteristics that are simpler to observe, such as age or place of residence. Besley and Kanbur (1990) posit that the marginal costs of eligibility determination become prohibitive if a program strives to achieve close-to-perfect targeting.

This hypothesis is largely refuted by the available cost data. In most cases, administrative costs represent a small portion of the total program budget, even for programs that differentiate eligibility or benefits by welfare level. The administrative costs associated with eligibility determination methods are a subset of that low total. This pattern is also observed for the cost of social registries relative to the social assistance programs they serve.

Calculating and comparing administrative costs is a bit tricky and thus requires some definitions. The administrative costs of a social assistance program include all the expenditures needed to design and implement the program—all the costs over and above the cost of the transfers but not including services that may also be provided (counseling, coaching, and training). Administrative costs are incurred by all social assistance programs, be they narrowly targeted or not. Such costs include the costs of planning and information systems, mechanisms for payments, grievance redress, audits, monitoring and evaluation, and so forth. In addition, narrowly targeted programs would incur other (somewhat higher) costs associated with eligibility determination and recertification. These are the marginal administrative costs associated with narrow targeting, which drive Besley and Kanbur's (1990) theoretical argument.

There is scant information on the level of administrative costs of social assistance programs in low- and middle-income countries because it is quite difficult to collect. Some of these costs are incurred primarily at the central level, others in frontline units, and others by third-party agencies (for example, payments). To get a comprehensive estimate of a program's administrative costs, the program administration or cost expert should collect all these elements from all the cost centers. Often, some of the program resources—such as staff or other operational systems—are shared across multiple programs; in these cases, the costs must be assigned to or shared

among the specific programs using different allocation keys. Moreover, depending on the maturity of the program, the level of administrative costs may decrease over time, as the fixed costs of design, investments in information systems, and monitoring and evaluation are spread over a larger caseload and years in service.[23] All these factors make the careful estimation of administrative costs a complex exercise, which is infrequently done for a subset of programs or even social registries (which usually support multiple programs). Therefore, this subsection relies on partial information that must be interpreted carefully.

Information on the composition of these administrative costs, including the marginal costs associated with narrower targeting, is rarer still. All programs, whether welfare targeted or not, incur expenditures to identify beneficiaries or transfer payments. For example, a child allowance for all children requires registering them in the information system, including information on the responsible adult with identity information, contact information, payment information, and so forth. For a welfare-targeted program, the marginal cost of targeting would be related to the information needed to measure, estimate, or rank the welfare of households (in community-based targeting, proxy means test, hybrid means test or means-tested programs) and update that periodically if needed. Thus, administrative costs per client may be higher for narrowly targeted programs, but overall costs may not. Figure 2.14 shows how the costs of a poverty-targeted child allowance might look versus a universal child allowance.

Three statistics are often used to report on the level of administrative costs, including the subset of costs associated with narrower targeting, across programs and countries: the cost per beneficiary served, the share of administrative costs in total program costs, and the cost-transfer ratio. The cost-transfer ratio measures the cost of making a one-unit transfer to a beneficiary. However, interpreting these statistics is not without problems. The cost per beneficiary needs to be converted from local currency into US dollars or another internationally used currency and will depend on the type of exchange rate used (official or based on purchasing power parity and for some countries, unofficial). The share of administrative costs in total program costs (cost of transfers plus administrative costs) and the cost-transfer ratio (a linear transformation of the former) are probably the simplest ways to look at these costs across programs. However, they are influenced by various program design parameters, such as the size (coverage) of the program, its maturity (pilot versus at-scale), the size of the transfer, the type of targeting method or methods used, the frequency of recertification, the feasibility and use of program data interoperability, the type of payment mechanism, and so forth. A program with a more generous benefit would score a smaller cost-transfer

Figure 2.14   Costs of Targeted versus Universal Child Allowance
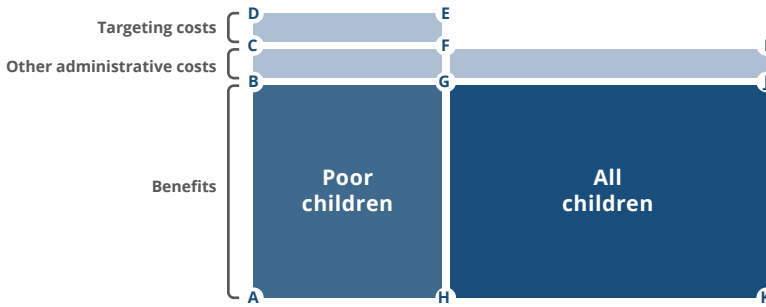
Source: Tesliuc et al. 2014

ratio or share of administrative costs indicator, all else remaining equal. (See Caldes, Coady, and Maluccio [2006] or Tesliuc et al. [2014] for more on these points.)

Empirical studies have found that the administrative costs associated with targeted programs represent a small share of total program costs. The few empirical studies on administrative costs suggest that most well-run social assistance programs operate with a modest share of administrative costs in the total program budget. Grosh et al. (2008) report a share of administrative costs averaging 5 to 10 percent for public works and cash transfer programs (conditional or not) and 22 percent for in-kind programs (across 55 social assistance programs). Tesliuc et al. (2014) report a range from 2 to 10 percent for a last-resort cash transfer program in the Europe and Central Asia region.[24] Schnitzer and Stoeffler (2020) similarly report a low share of administrative costs in total program costs of 0.4 to 5.5 percent for programs in Burkina Faso, Chad, and Niger, depending on the approach used by the administrators.[25] Rosas, Zaldivar, and Pinzon-Caicedo (2016) report a cost of 7.7 percent during the first phase of implementation of the Tanzania Productive Safety Net Program. Jamaica's Advancement Through Health and Education program registered a slightly higher share of administrative costs of about 11 percent for 2018, which includes, in addition to the delivery of conditional cash transfers, case management services and the cost of associated social workers (World Bank 2019).

The costs of social registries, which are used by multiple programs per country to determine eligibility, are also small compared with total program costs. A significant number of developing countries have established social registries to collect information, in office or in the field, to establish eligibility for social programs. Their costs become the new pertinent way to

measure the administrative costs of targeting in many settings. A literature review of the costs of large-scale social registries in middle-income countries, which support multiple targeted programs, found that these costs range between US$1 and US$3 per household in most countries, or less than 2 percent of the value of benefits channeled through the targeting system. More specifically, the cost per registered household was US$1.4 dollars in Pakistan; around US$2.5 in Bangladesh, Indonesia, and the Philippines (Packard et al. 2019); US$1.3 in Turkey; US$1.27 in Colombia for System for the Selection of Beneficiaries for Social Programs (SISBEN) IV; US$2.25–US$2.50 in Colombia for SISBEN I, II, and III (Departamento Nacional de Planeación database; Leite et al. 2017); and US$2.06 in Brazil for the Cadastro Único 7 until 2013 (Leite et al. 2017). Castaneda and Lindert (2005) report different statistics for the share of the program budgets that were targeted through the social registry: Brazil, 1.4 percent; Chile, 1.3 percent; Colombia, 0.9 percent, Costa Rica, 0.5 percent; and Mexico, 0.7 percent. Hanna and Olken (2018) cite values of 0.8 and 1.7 percent for total transfer costs going to the costs of the social registries in Indonesia and Peru, respectively. Lindert et al. (2018) show that for Malawi's National Social Support Programme Phase 2, the cost was US$1.74 per household. A recent review of the costs of 10 social registries[26] confirms that the administrative costs associated with finer targeting are low, within 1 to 3 percent of the value of the annual transfers (figure 2.15 reports the cost-transfer ratios for the social registries). These costs are in the same range as those reported in other reviews, based on overlapping country coverage (see Devereux et al. 2017; Kidd, Athias, and Mohamud 2021).

Not surprisingly, unit costs are higher for social registries in low-income countries that are in an incipient or partial rollout phase,[27] but they continue to remain small compared with the value of the transfers. For example, the cost-transfer ratios of the incipient social registries established in the Republic of Congo and Mali,[28] at 7 and 8 percent (figure 2.15), are higher than the cost-transfer ratios for social registries that have achieved national scale. The Republic of Congo and Mali have established registries in an environment of paucity of other administrative data sources (which could have reduced the cost of data collection and verification) and are only in the initial phase of their expansion (not yet benefiting from the economies of scale of a social registry with larger coverage) (see also annex 2D).

Increasingly, programs or social registries collect less information directly from the beneficiaries, while using more information from other administrative databases through interoperability and cross-matching with other databases held by the government, such as income tax, social security contributions, registration of land or automobiles, passports and payments to government-operated utilities, and so forth. In such cases, the costs of

Figure 2.15   Ratio of the Unit Cost of a Social Registry versus the Annual Benefit of the Largest Program Served
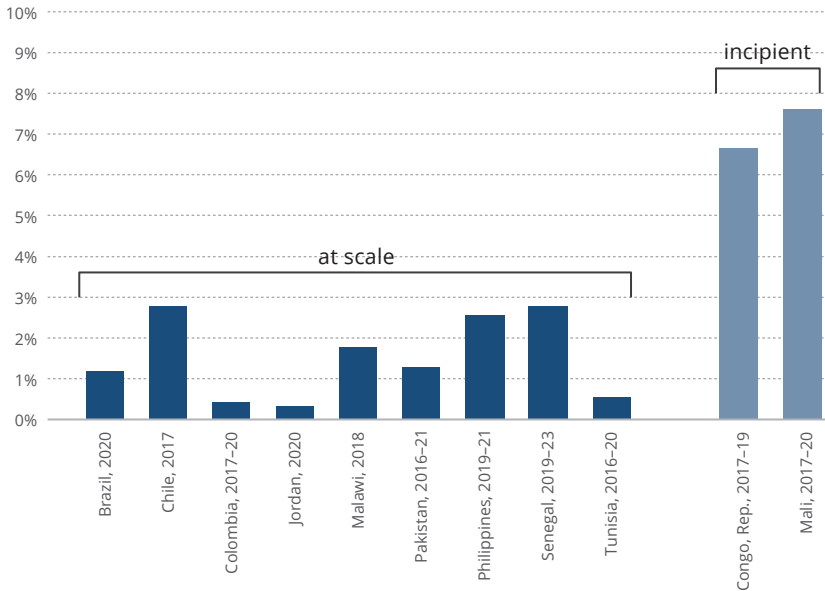
Source: Original compilation for this publication based on Annex 2D.

running those other databases, whose main functions are in their home ministries, are not considered part of the administrative costs of the social assistance program. However, it may take some investment on the part of the social registry to be able to use such data, and it may increase the frequency of updates so that when the overall costs associated with interoperability are included in the social registry costs, there is an increase in both administrative cost and the efficiency or productivity of the system.

For example, in Turkey, the estimated development cost of the Integrated Social Assistance System (ISAS) was US$13.1 million and it was built between 2010 and 2015, reaching about 40 percent of the population in 2015.[29] Over 2010–15, ISAS served 43 million people with a unit cost over five years of US$0.3. The investment in ISAS allowed the rationalization of social assistance benefits, by identifying duplications of about 10 percent. Making processes electronic also saved costs by reducing paper and staff time; the government now processes approximately 2.3 million fewer documents per month. In addition to this, processing time has been significantly reduced. For example, the time from application to decision for

regular social assistance programs has been reduced by approximately 20 percent, and the time from application to disbursement to beneficiaries of the disability and old-age pension programs provided under Law No. 2022 dropped from 1.5 years to one month.[30]

In Brazil, since 2016, Cadastro Único communicates with 10 other information systems to verify inconsistencies between the declared income and the individual information available in the other information systems. As a result, Brazil can regularly update its caseload of beneficiaries and save millions of dollars in fraud. For example, in 2019,[31] it was estimated that the Brazilian government had removed about 1.3 million people from the Bolsa Familia program, generating savings of about R$1.4 billion (US$350 million in 2019) due to the interoperability and cross-verification process. The increase in complexity and functionalities has pushed the average cost per household to US$6.7, compared with US$2.0 during 2010–13; however, in relative terms, the cost of Cadastro Único represents only about 1 percent of the annual transfer cost of the Bolsa Familia program (and is used for eligibility determination in a score of other programs).

Over the next three to four years, Colombia is expected to invest significantly in improving the interoperability and dynamism of SISBEN, increasing the unit cost of application from US$2.25–US$2.50 to about US$6, matching the unit costs in Brazil and Chile after their investment in interoperability was completed.

Administrative costs can be viewed as the investment needed to produce good outcomes (for example, to improve delivery systems; see chapter 4), and there is evidence that a somewhat higher share of "marginal targeting costs" in expenditures can improve targeting accuracy. Tesliuc et al. (2014) generate one of the more thorough cross-country comparisons of administrative costs in total and by their various functions. They find a strong correlation between the cost-transfer ratio of last resort income support programs in Europe and Central Asia and the share of benefits reaching the poorest quintile (figure 2.16). There is a range of optimal investment in program administration: lower spending would result in large errors and hence diminished cost-efficiency and effectiveness, and after a certain point, higher administrative costs indicate waste. Programs should finance enough of these costs, especially when they are the critical factor determining the effectiveness/efficiency of the cash or in-kind transfers.

The administrative effort and political will put into developing poverty-targeting and/or social registry systems has been substantial, but the choices made and scale have kept the unit costs low relative to the benefits channeled. In general, social protection delivery systems in developing countries could benefit from more investment in administrative functions. The experience has been that offices are few and often far from beneficiaries, with underdeveloped information systems or too few staff, and with

Figure 2.16 **Share of Program Benefits Reaching the Poorest Quintile and the Cost-Transfer Ratio**
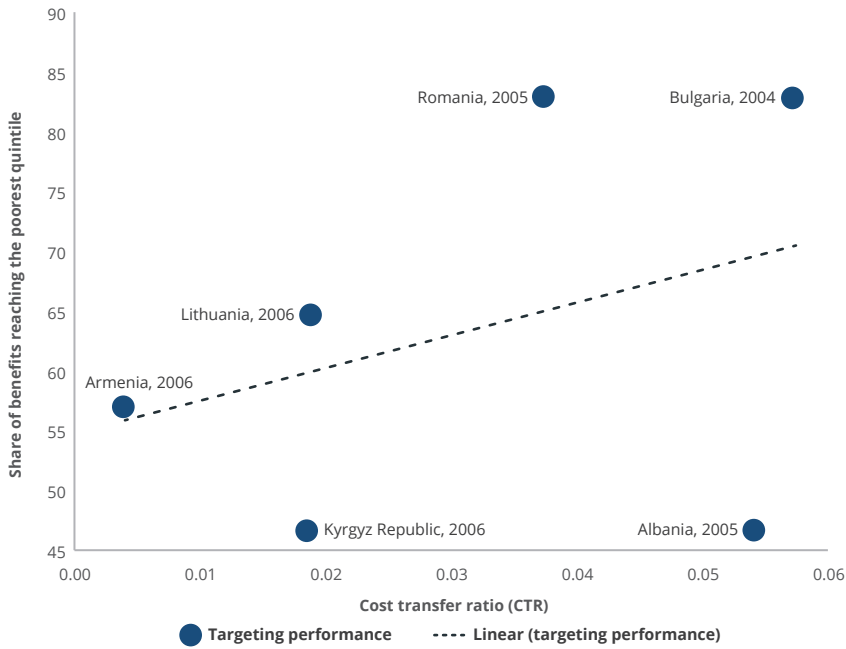
Source: Tesliuc et al. 2014

inadequate abilities to do outreach to the intended population or address grievances. Similarly, their "virtual portal" systems may be underdeveloped and suboptimally user-friendly (see chapter 4; Lindert et al. 2020). Scrimping on administration of social protection delivery systems may just shift costs to clients, raising their transaction costs and contributing to stigma, and increase the level of errors in the program. Given that transfer costs often account for 90 to 95 percent of total program costs, it would seem that programs should err on the side of further investment in delivery systems instead of scrimping to lower administrative costs. Both transaction costs and stigma are issues that can be reduced with good human-centered design, keeping an eye on client experience, investing in administrative systems that facilitate easy access, and so forth. It is likely that technology can help keep administrative costs manageable, as new ways of identifying and paying beneficiaries are developed and operational systems become less fragmented and are shared across different programs.

Transaction Costs

Transaction costs of various forms can reduce the value of participating in programs and sometimes exclude people altogether. Finding out about a program, filling out forms, supplying proof of identity and other required documents, being interviewed, and following up can all take time, may require bus fares or fees for obtaining documents, and so forth. The problem is intuitive to understand, and a few studies show that for at least some members of the target population, the barriers can be significant. Daigneault, Jacob, and Tereraho (2012) find that basic information and the characteristics of the claiming process are the two most commonly cited factors in their study of take-up of benefits in a few Organisation for Economic Co-operation and Development (OECD) countries. Delaney and Jehoma (2016) find that about 18 percent of income-eligible children do not receive the South African Child Support Grant for such reasons, although the problem is not restricted to targeted programs. ILO (2014) shows that participation in Namibia's "universal" (age- but not poverty-targeted) social pension is about 92 percent of those over age 60 years. Thinkthrough (2021) documents that the transaction costs for Nepal's universal child allowance are a barrier to participation, especially for households in remote areas who must travel to reach pertinent administrative services, compounded by the incomplete coverage of birth registration, the program's relatively complex administrative procedures, and the relatively low value of the transfer.

Although there is still far to go, there is a great deal of know-how and many examples of its utilization to show that social assistance programs could effectively tackle issues of transaction costs, which are discussed in more depth in chapter 4. An increasing number of countries are working to ensure that residents or citizens have identity documents, which is a common stumbling block, with almost every country in Africa and Asia having introduced an electronic identification (eID) or intending to do so in the near future. Social registries to serve multiple programs are being built in many countries (Leite et al. 2017). The increasing use of digital payment systems, especially those that allow multiple options of service providers, is reducing the transaction costs of collecting benefits, which is important, as it is a recurrent rather than one-off transaction. In addition to developing these basic systems, many countries have initiatives for "active outreach" as part of their social protection activities. In Brazil, for example, an active outreach strategy for the social registry was initiated in 2011 with the tagline *"Conhecer para Incluir"* (to know so as to include). The outreach effort was intense until 2014 and included media outreach and door-to-door efforts in target areas from slums to jungles. About 1.5 million new families were added to the national social registry, which is

used for 30 poverty-targeted programs, principal among them the Bolsa Familia conditional cash transfer program. Of the families added, over a million were traditional groups (indigenous, quilombolas [residents of Afro-Brazilian communities], or riverine populations) that are highly vulnerable and often underserved.

Social Costs

Programs that differentiate eligibility by welfare may lead to jealousies or ill-will in a community between recipients and nonrecipients. The qualitative research that it takes to detect such effects is not as common as the more quantitative impact evaluations. Qualitative studies are often tied to the early phases of program implementation when the idea of the program is new in the researched communities and the implementation bugs have not yet been ironed out. For example, Della Guardia, Lake, and Schnitzer (forthcoming) investigated the effects of Chad's pilot cash program, which was initiated in 2016, geographically targeted to the poorest rural areas, with caseloads allocated so that about 40 percent of the people in each included village would benefit, and a proxy means test to select them. The program increased the participants' consumption and investment (Kandpal, Schnitzer, and Daye 2020). There were positive local spillover effects—some recipients shared their transfers directly with family or neighbors, some helped create community infrastructure, and there was a positive effect on local small businesses and the market for day labor as the beneficiaries spent their transfers and invested in their household enterprises. But there were jealousies as well, and recipients reported that nonrecipients were sometimes rude, jealous, or angry, and sometimes they took actions that were economically punitive, for example, charging higher prices in local commerce, refusing to give full change in transactions, or refusing to repay credit. This level of backlash is more marked than in some other reports, which include some friction, gossip, and repercussions for community labor that is not directly associated with the program but not economic retaliation per se.

Jones, Vargas, and Villar (2007) conducted field work in the early days of Peru's Juntos conditional cash transfer program, which geographically targets rural villages with beneficiaries selected through a proxy means test and a final community validation phase. The program has a record of positive impact evaluations along the usual dimensions for quantitative impact evaluations of conditional cash transfer programs, such as increased consumption; increased school enrollment, attendance, and grade progression; increased use of health care; and mildly better nutrition outcomes (Jaramillo and Sánchez 2011; Perova and Vakis 2009; Perova and Vakis 2012). But in their qualitative field work, Jones, Vargas, and Villar (2007) found some

issues of jealousies, including nonrecipient families who were jealous that children in recipient families had better school uniforms and shoes; that nonrecipient youth were more reluctant to contribute to school chores, saying the recipient youth should do them; and that some mothers felt cohesion in community activities had lessened.

There are similar accounts from the outset of Lesotho's Child Grant Programme (OPM 2014) and Mexico's *PROGRESA* program (Adato 2000). The resentments in all cases were in village settings where people knew each other; these feelings may not carry through so much to more urban settings (where an increasing share of the world lives and the new wave of social assistance programs in Africa are located). But they are tied up in the issues inherent in targeting; that is, the rationing of spots in a program in which people want to participate and the difficulty of drawing clear distinctions among the more and less needy in fairly homogeneous poor places.

Stigma is somewhat the inverse of jealousy, and it can be thought of as the psychic version of transaction costs and may be tied up in some of the same processes. Program recipients may have to identify themselves publicly or semipublicly as in need of help. This may entail queuing at social service assistance centers or payment collection points; having their name on a list of aid recipients posted as part of transparency initiatives; or being cross-questioned on income and expenditures, work, school attendance, or other behaviors. These experiences can feel demeaning and all the more so if the public who might witness, or particularly the staff involved in program administration, convey through word or gesture negative judgments about the program claimants. These may be putative behaviors (they are lazy, dirty, cheat, and so forth) or group identities they may hold (educational background, ethnicity, religion, native language, or migration status). Full-scale research in this area is relatively rare, although the problem seems to be common. Baumberg (2016) gives an overview for the UK, Yang et al. (2019) provide a literature review around child benefits, and Gubrium and Pellissery (2016) review a cross-section of programs. In OECD countries, Daigneault, Jacob, and Tereraho (2012) find that stigma is the sixth most common reason for non-take-up of benefits, cited in 22 percent of the studies reviewed. Wright et al. (2015) find that in 26 of 30 focus groups in their research on the dignity of claimants of South Africa's Child Grant, claimants found issues in which the application process affected their dignity, including long queues, having to negotiate the application process, and being treated disrespectfully by government officials.

The issues of stigma and shame are intertwined, but they are not identical. Shame can come from poverty itself and the way it can constrain self-esteem or the ability to engage in social roles. Some people report that the

recognition of their needs in the social assistance program is in itself an affirmation and alleviates rather than induces shame (see Gubrium and Pellissery 2016; Roelen 2017; Yang et al. 2019). Granlund and Hochfeld (2020) show that South Africa's Child Grant has largely positive effects on the dignity and autonomy of the caregiver recipients. A growing body of impact evaluations shows how receipt of social assistance can improve psychosocial well-being, including through allowing participants to become more engaged in social networks (for example, Attah et al. [2016] provides a review of cash transfers in five African countries). The additional resources from the transfer benefits may reduce shame, although the process of claiming the benefits may be more or less stigmatizing, depending on administrative processes, program design, and context (see Roelen [2020] for a review of the interaction).

The connotations and framing of a program can seemingly influence how people feel about it. In the United States, for example, Pell Grants are means-tested federal aid for the costs of college. They are generally not viewed as stigmatizing, presumably because the very act of going to college is a triumph and a step on a journey that is likely to lead to a good job. The thresholds for Pell Grants are set high, so they are better thought of as affluence tested than poverty targeted. Many countries try to associate their programs with positive messages through their names. For example, Mexico's cash transfer program was first called PROGRESA, an acronym for Progressing through Education and Health, then Oportunidades (opportunities), and then Prospera (prosper). Indonesia's conditional cash transfer program is called PKH (the Family Hope Program), Peru's is Juntos (Together), and Jamaica's PATH stands for Program of Advancement Through Health and Education. Mali's program is called Jigisemejiri (Tree of Hope), and the Philippines' 4Ps is the Bridging Program for the Filipino Family.

The issues of exclusion caused by transaction costs and stigma are partly the result of delivery systems that are insufficiently developed and/or inattentive to clients' needs. As shown in chapter 4, there is much that can be done to ameliorate these problems if there is political will and sufficient funding for administrative costs. In addition to all the good that can be done through making transactions in social assistance service centers convenient and nontraumatizing, the move to digital may also help to reduce stigma. The more transactions are private, the fewer will be the occasions in which people will be treated badly, especially in front of their community. For example, the move to payment via debit cards has been welcomed by beneficiaries of programs such as the US food stamp program and Brazil's Bolsa Familia cash transfer, because their cards made them look like better-off consumers with a regular bank debit or credit card (Oliveira et al. 2018).

Summary

This short benchmarking has underscored important features of empirical outcomes from social assistance practice in developing countries:

- *Details of data matter.* This chapter exploited the strength of data for a large number of countries and standardized methods, but it was somewhat hampered by the limitations of grouping program observations within harmonized program types and relatively sparse coverage of low-income countries.
- *The degree of variation in every indicator examined is quite remarkable.*
- *Regularities are nonetheless present and in expected directions:*

  o Coverage by social assistance is 36 percent for the world as a whole and 54 percent for the poorest quintile.
  o The range of coverage among countries is quite broad, from virtually none to virtually all of the poorest quintile covered, but there are truncations of the range at the extremes of the country income groups. None of the low-income countries manages to cover more than about half the poorest quintile, and none of the World Bank client high-income countries covers less than half the poorest quintile. In low-income countries, on average, only 17 percent of the people in the poorest quintile are covered by social assistance, whereas the figure is 77 percent for high-income countries.
  o Incidence is progressive, but incidence graphs aggregated across countries by program type look more like a mildly sloped hill than a sharp step function. In the program aggregations, the share of beneficiaries in the poorest quintile ranges from 30 to about 50 percent, depending on program type.
  o There is even more variability at the level of the individual country and program type. Unconditional cash transfers are the program category with the most observations and the most variability. Between 6 and 73 percent of the beneficiaries of unconditional cash transfers are in the poorest quintile of the population.
  o Scoring well on incidence and coverage is associated with the degree to which the program lowers the poverty gap, but incidence much more strongly so. Very large programs tend to have high coverage of the poor but less progressive incidence. An additional 10 percent share to the poorest quintile means the poverty gap falls by an extra $0.10 per $1 spent.
  o To produce the strongest impacts on poverty requires both good coverage of the poorest and a meaningful level of benefits. To do it in a fiscally affordable way also requires relatively progressive incidence.

- *Using multiple measures of outcomes to consider a program is important.* Unidimensional approaches can give only partial views, at the risk of missing important factors.
- *It is important to consider costs.*

  ○ Concerns about labor disincentives are intuitive and much discussed in the more theory-based literature, but social assistance program designs in developing countries often avoid the features that might be most fraught with disincentives. A growing and robust impact evaluation literature shows that, in general, labor disincentives have not been a big issue; indeed, the programs studied sometimes increase work, formality, or earnings.

  ○ Another often-cited argument against differentiating eligibility or benefits by welfare level is that the administrative costs of doing so can be high. Again, this hypothesis is largely refuted by the available cost data. The costs of large-scale social registries in middle-income countries, which support multiple targeted programs, range between US$1 and US$3 per household in most countries, or 1–3 percent of the value of benefits channeled through the targeting system. Relative costs are higher in low-income countries with more nascent systems, as they have not yet "amortized" the startup costs or reached large scale, on the order of 7–8 percent.

  ○ Transaction costs are also nonzero, but their most worrisome aspect is in the errors of exclusion that can result, and those are accounted for already in the coverage/incidence counts.

  ○ The stigma or reduction in social cohesion that may result from focusing benefits on the poorer and the loss of political support (discussed in the chapter 1) are the most difficult factors to quantify and weigh in the ledger of pros and cons.

Although the presence of imperfections and costs may be inherent in targeting, the present magnitude of them is not. Current practice shows great variability, suggesting that a great deal of improvement is possible if programs emulate the best practices of others. Countries need to doggedly build the capacities that will allow successful social policy, including the systems that support differentiation of eligibility or benefits where that is a tactic taken. Over the past two decades, there have been many examples of such capacity being built. But there are also examples where once rapid advance has slowed and a few cases of retrocession. The next chapters take up the how-to of focusing resources on the lower end of the welfare spectrum, highlighting progress and pitfalls.

Annex 2A: List of ASPIRE Household Surveys Used in the Analysis

Table 2A.1: **List of ASPIRE Household Surveys Used in the Analysis**

Country/ Economy	Country code	Year	Region	Country income group	Survey name
Argentina	ARG	2018	LAC	UMIC	Encuesta Permanente de Hogares Continua
Armenia	ARM	2018	ECA	UMIC	Armenia—Integrated Living Conditions Survey
Azerbaijan	AZE	2015	ECA	UMIC	Azerbaijan Monitoring Survey of Social Welfare 2015
Bangladesh	BGD	2016	SAR	LMIC	Household Income and Expenditure Survey 2016
Belarus	BLR	2017	ECA	UMIC	Household Sample Survey
Bolivia	BOL	2018	LAC	LMIC	Encuesta de Hogares
Bosnia and Herzegovina	BIH	2015	ECA	UMIC	Household Budget Survey
Botswana	BWA	2015	SSA	UMIC	Botswana Multi-Topic Household Survey
Brazil	BRA	2018	LAC	UMIC	Pesquisa Nacional por Amostra de Domicílios Contínua
Burkina Faso	BFA	2014	SSA	LIC	Enquete Multisectorielle Continue
Cameroon	CMR	2014	SSA	LMIC	Quatrième Enquête Camerounaise Auprès des Ménages 2014 (ECAM4)
Chile	CHL	2017	LAC	HIC	Encuesta de Caracterización Socioeconómica Nacional (CASEN)
Colombia	COL	2018	LAC	UMIC	Gran Encuesta Integrada de Hogares 2018
Costa Rica	CRI	2018	LAC	UMIC	Encuesta Nacional de Hogares (ENAHO)
Côte d'Ivoire	CIV	2015	SSA	LMIC	Enquête sur le Niveau de Vie des Ménages de Côte d'Ivoire 2015
Croatia	HRV	2014	ECA	HIC	Household Budget Survey
Dominican Republic	DOM	2018	LAC	UMIC	Encuesta Nacional de Fuerza de Trabajo
Ecuador	ECU	2018	LAC	UMIC	Encuesta Nacional de Empleo Desempleo y Subempleo
El Salvador	SLV	2018	LAC	LMIC	Encuesta de Hogares de Propósitos Múltiples
Eswatini	SWZ	2016	SSA	LMIC	Household Income and Expenditure Survey
Ethiopia	ETH	2018	SSA	LIC	Ethiopia Socioeconomic Survey (ESS)
Gambia, The	GMB	2015	SSA	LIC	Integrated Household Survey

Table 2A.1 *(continued)*

Country/ Economy	Country code	Year	Region	Country income group	Survey name
Georgia	GEO	2018	ECA	UMIC	Household Integrated Survey
Ghana	GHA	2016	SSA	LMIC	Living Standards Survey VII 2016–17
Guatemala	GTM	2014	LAC	UMIC	Encuesta Nacional de Condiciones de Vida
Honduras	HND	2017	LAC	LMIC	Encuesta Permanente de Hogares de Propósitos Múltiples
Indonesia	IDN	2018	EAP	UMIC	Survei Sosial Ekonomi Nasional 2018, Maret (SUSENAS)
Jamaica	JAM	2017	LAC	UMIC	Survey of Living Conditions
Kazakhstan	KAZ	2017	ECA	UMIC	Household Budget Survey
Kenya	KEN	2015	SSA	LMIC	Integrated Household Budget Survey 2015–16
Lesotho	LSO	2017	SSA	LMIC	Household Budget Survey
Liberia	LBR	2016	SSA	LIC	Household Income and Expenditure Survey 2016
Malawi	MWI	2016	SSA	LIC	Fourth Integrated Household Panel Survey 2016
Malaysia	MYS	2016	EAP	UMIC	Household Income Survey
Maldives	MDV	2016	SAR	UMIC	Household Income and Expenditure Survey 2016
Mauritania	MRT	2014	SSA	LMIC	Enquete Permanente Sur Les Conditions De Vie Des Menages
Mauritius	MUS	2017	SSA	HIC	Household Budget Survey
Mexico	MEX	2018	LAC	UMIC	Encuesta Nacional de Ingresos y Gastos de los Hogares
Moldova	MDA	2018	ECA	LMIC	Household Budget Survey
Mongolia	MNG	2016	EAP	LMIC	Household Socio-Economic Survey
Montenegro	MNE	2014	ECA	UMIC	Household Budget Survey
Mozambique	MOZ	2014	SSA	LIC	Inquérito Sobre Orçamento Familiar 2014-2015
Myanmar	MMR	2017	EAP	LMIC	Myanmar Living Conditions Survey
Namibia	NAM	2015	SSA	UMIC	National Household Income and Expenditure Survey 2015–16
Nicaragua	NIC	2014	LAC	LMIC	Encuesta Nacional de Hogares sobre Medición de Nivel de Vida
Niger	NER	2014	SSA	LIC	Enquête Nationale sur les Conditions de Vie des Ménages et l'Agriculture
Nigeria	NGA	2018	SSA	LMIC	Nigeria Living Standards Survey
Pakistan	PAK	2018	SAR	LMIC	Social and Living Standards Measurement (PSLM_HIES) 2018–19

Table 2A.1 *(continued)*

Country/ Economy	Country code	Year	Region	Country income group	Survey name
Panama	PAN	2018	LAC	HIC	Encuesta de Mercado Laboral
Paraguay	PRY	2018	LAC	UMIC	Encuesta Permanente de Hogares
Peru	PER	2017	LAC	UMIC	Encuesta Nacional de Hogares
Philippines	PHL	2015	EAP	LMIC	Family Income and Expenditure Survey 2015–16
Poland	POL	2015	ECA	HIC	Household Budget Survey
Romania	ROU	2016	ECA	HIC	Household Budget Survey
Russian Federation	RUS	2017	ECA	UMIC	Statistical Survey of Income and Participation in Social Programs
Serbia	SRB	2015	ECA	UMIC	Household Budget Survey
Sierra Leone	SLE	2018	SSA	LIC	Integrated Household Survey
South Africa	ZAF	2014	SSA	UMIC	Living Conditions Survey 2014/15
Sri Lanka	LKA	2016	SAR	LMIC	Household Income and Expenditure Survey 2016
Tanzania	TZA	2014	SSA	LMIC	LSMS—National Panel Survey 2014–15—Wave 4
Thailand	THA	2018	EAP	UMIC	Household Socio-Economic Survey
Turkey	TUR	2018	ECA	UMIC	Household Income and Consumption Expenditures Survey 2018
Uganda	UGA	2016	SSA	LIC	National Household Survey 2016–17
Ukraine	UKR	2016	ECA	LMIC	Household Living Conditions Survey
Uruguay	URY	2018	LAC	HIC	Encuesta Continua de Hogares
Uzbekistan	UZB	2018	ECA	LMIC	Listen to Citizen of Uzbekistan (L2CU)
Vietnam	VNM	2014	EAP	LMIC	Household Living Standard Survey
West Bank and Gaza	PSE	2016	MNA	LMIC	Expenditure and Consumption Survey 2016–17
Zambia	ZMB	2015	SSA	LMIC	Living Conditions Monitoring Survey VII 2015
Zimbabwe	ZWE	2019	SSA	LMIC	Mini Poverty, Income, Consumption and Expenditure Survey 2019

Source: ASPIRE.

Note: The sample size is 70 countries. EAP = East Asia and Pacific; ECA = Europe and Central Asia; HIC = high-income country; LAC = Latin America and the Caribbean; LIC = low-income country; LMIC = lower-middle-income country; MNA = Middle East and North Africa; SAR = South Asia; SSA = Sub-Saharan Africa; UMIC = upper-middle-income country.

Annex 2B: Distribution of Social Assistance Beneficiaries, by Program Type

Figure 2B.1 **Distribution of Conditional Cash Transfer Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare**
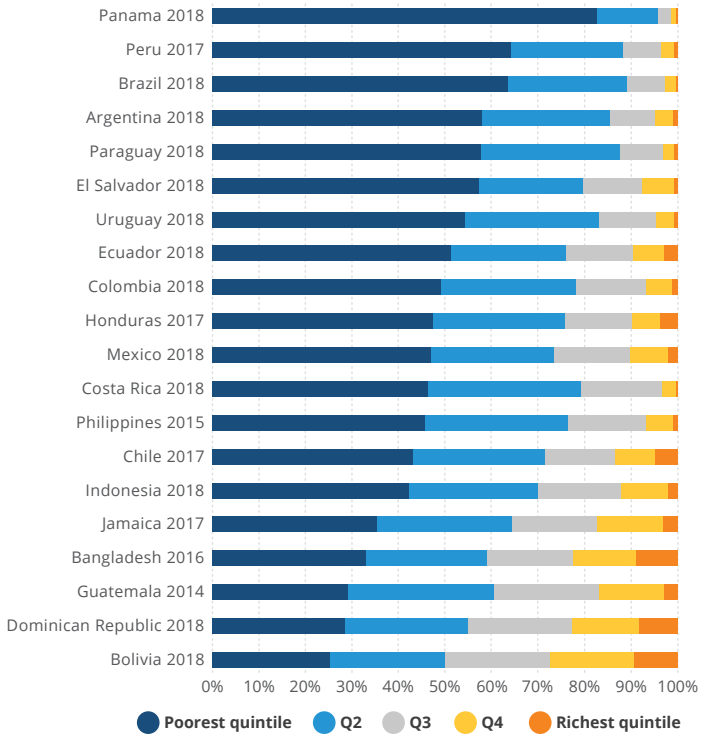
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: Total (n = 20), Latin America and the Caribbean (n = 17), East Asia and Pacific (n = 2), South Asia (n = 1). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2B.2   Distribution of Social Pension Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare
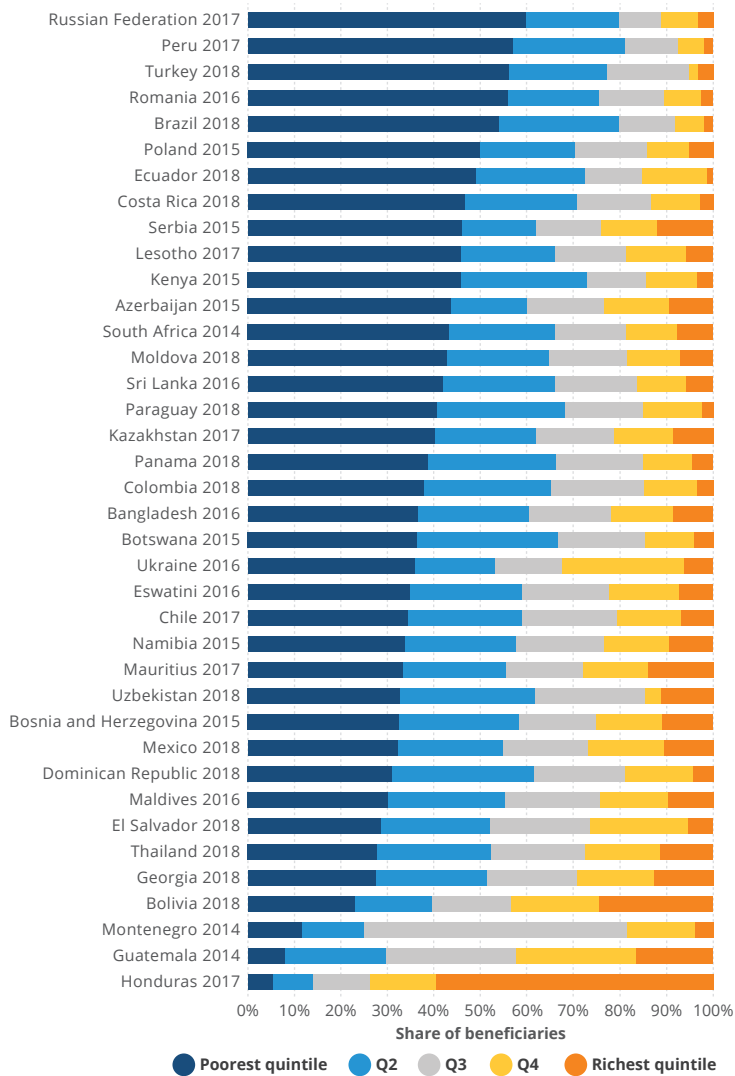
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: Total (n = 38), Europe and Central Asia (n = 13), Latin America and the Caribbean (n = 14), Sub-Saharan Africa (n = 7), South Asia (n = 3), East Asia and Pacific (n = 1). Social pensions include any of the following: noncontributory old-age, disability, and survivor pensions. Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2B.3 Distribution of In-Kind Transfer Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare
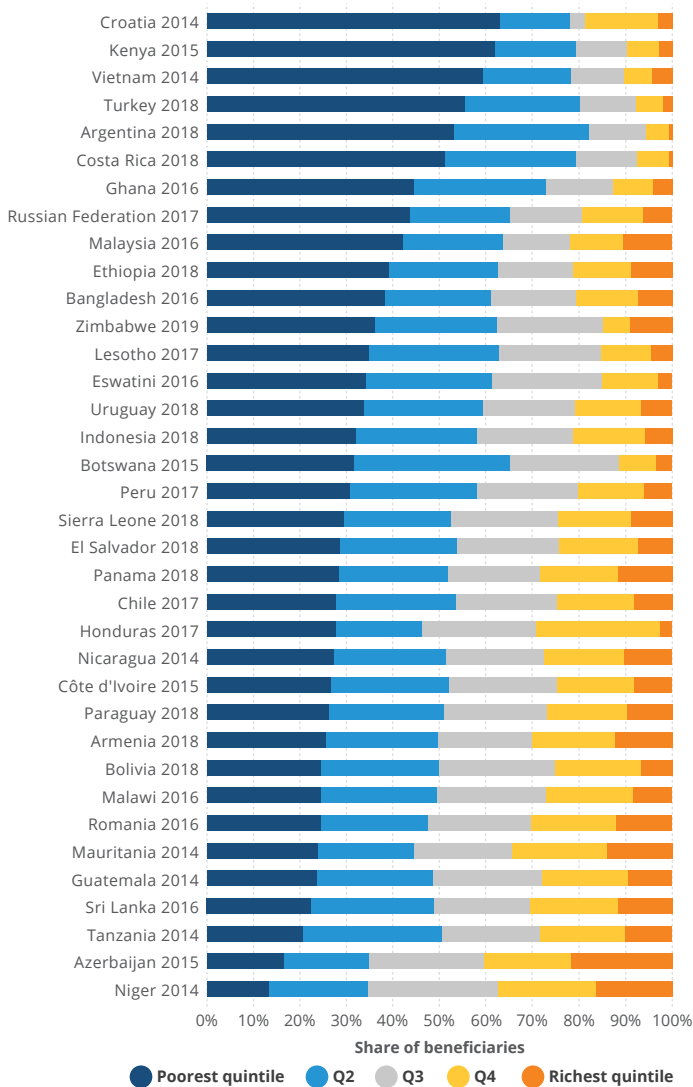
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: Total (n = 36), Sub-Saharan Africa (n = 13), Latin America and the Caribbean (n = 12), Europe and Central Asia (n = 6), East Asia and Pacific (n = 3), South Asia (n = 2). In-kind transfers include any of the following: food aid, agricultural inputs, clothes, school supplies, and building materials. Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2B.4 Distribution of Public Works Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare
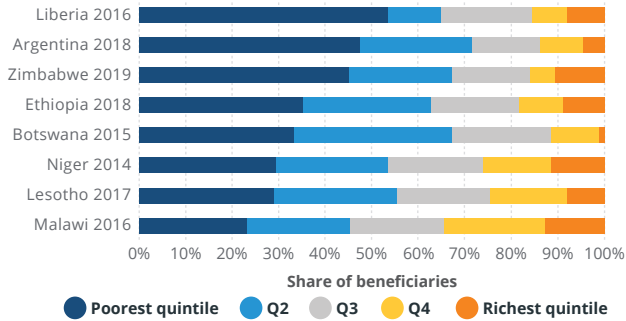
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: Total (n = 8), Sub-Saharan Africa (n = 7), Latin America and the Caribbean (n = 1). Public works programs include cash-for-work and food-for-work programs (including food for training and for assets). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2B.5 **Distribution of Fee Waiver and Targeted Subsidy Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare**
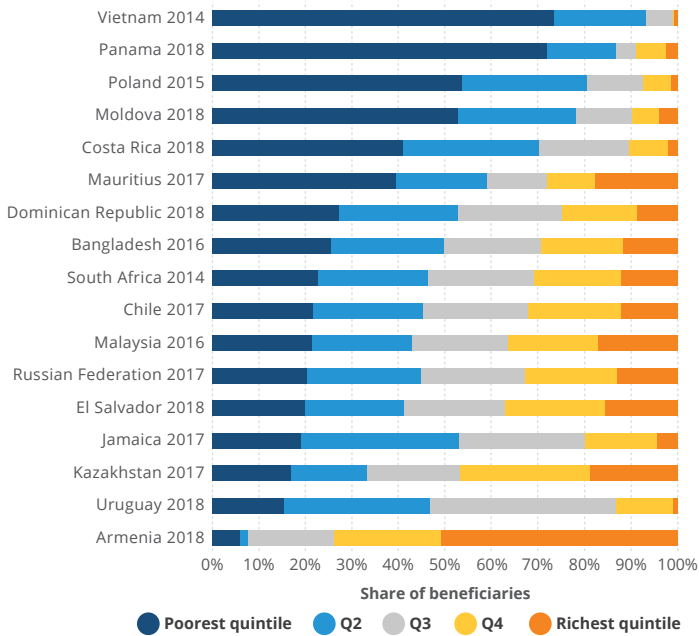
Source: ASPIRE database, www.worldbank.org/aspire.

Note: Number of countries: Total (n = 17), Latin America and the Caribbean (n = 7), Europe and Central Asia (n = 5), Sub-Saharan Africa (n = 2), East Asia and Pacific (n = 2), South Asia (n = 1). Fee waivers and targeted subsidies include any of the following: energy products, education, utilities, housing or transportation fees waivers to specific households, or discounted below the market cost. They do not include health benefits/subsidies. Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2B.6 Distribution of School Feeding Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare
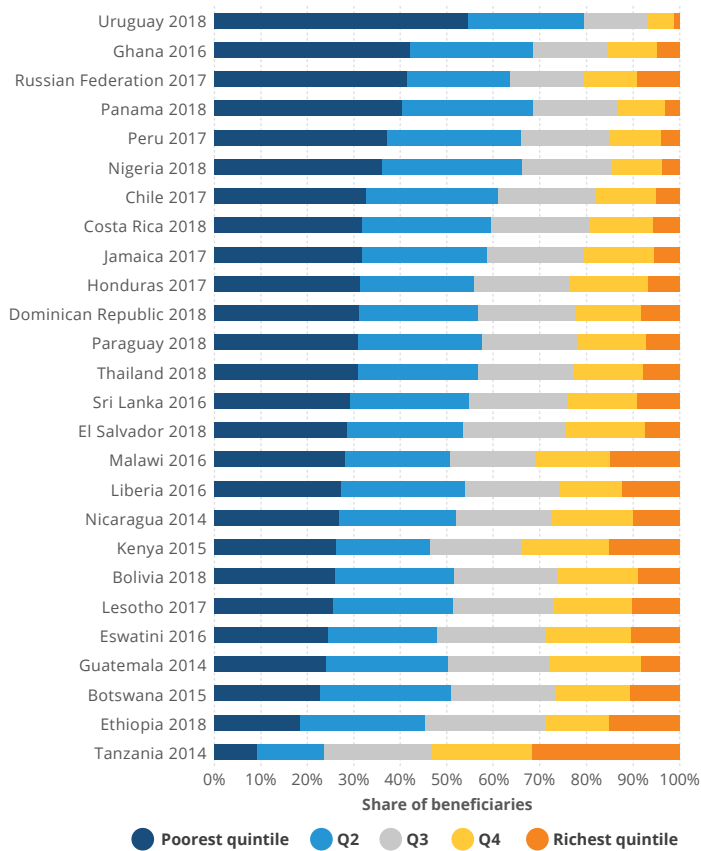
Note: The number of countries per region is as follows: Total (n = 26), Latin America and the Caribbean (n = 13), Sub-Saharan Africa (n = 10), Europe and Central Asia (n = 1), East Asia and Pacific (n = 1), South Asia (n = 1). School feeding programs encompass any type of meals or food items provided at school. Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2B.7   **Distribution of Other Social Assistance Beneficiaries, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare**
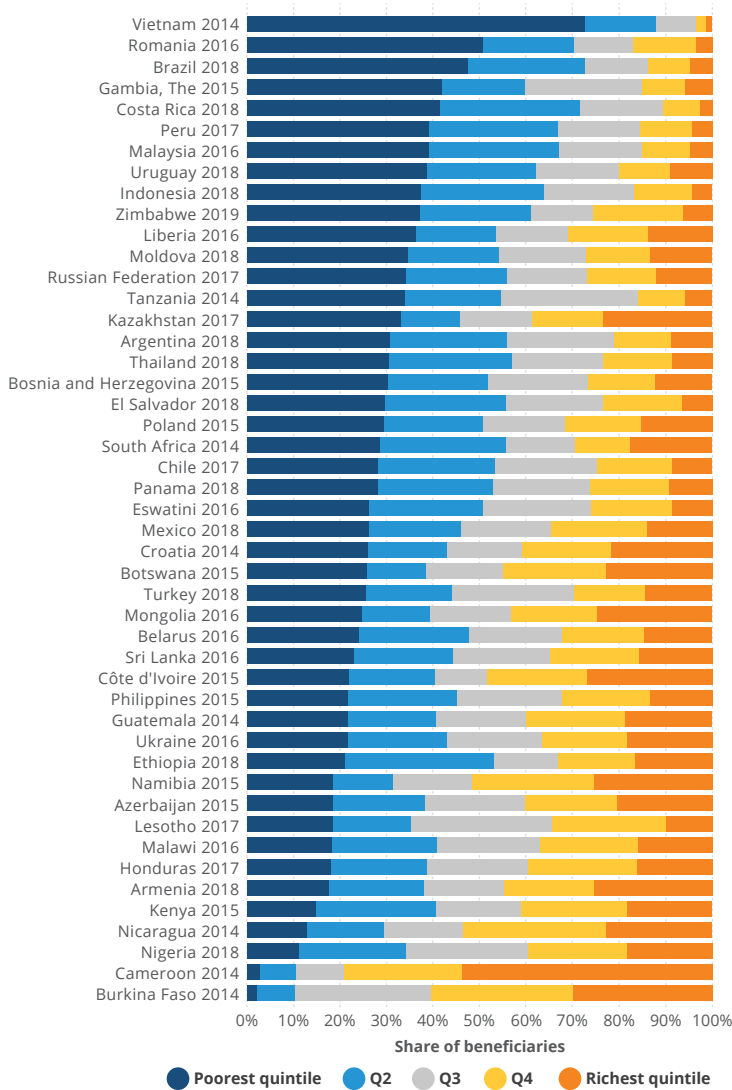
Source: ASPIRE database, www.worldbank.org/aspire.

Note: The number of countries per region is as follows: Total (n = 47), Sub-Saharan Africa (n = 16), Europe and Central Asia (n = 12), Latin America and the Caribbean (n = 12), East Asia and Pacific (n = 6), South Asia (n = 1). Other social assistance includes the following: scholarships/education benefits; social care services and other miscellaneous programs. Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Annex 2C: Coverage and Distribution of Social Assistance Beneficiaries

Figure 2C.1   Coverage and Beneficiary Incidence of Unconditional Cash Transfer Programs in South Africa, as Captured in the ASPIRE Household Surveys, by Quintile of Pretransfer Welfare

a. Coverage of unconditional cash transfer programs

b. Incidence of unconditional cash transfer beneficiaries

Legend: All conditional cash transfers ● ●   SENAPAN ● ●   Red de oportunidades ● ●
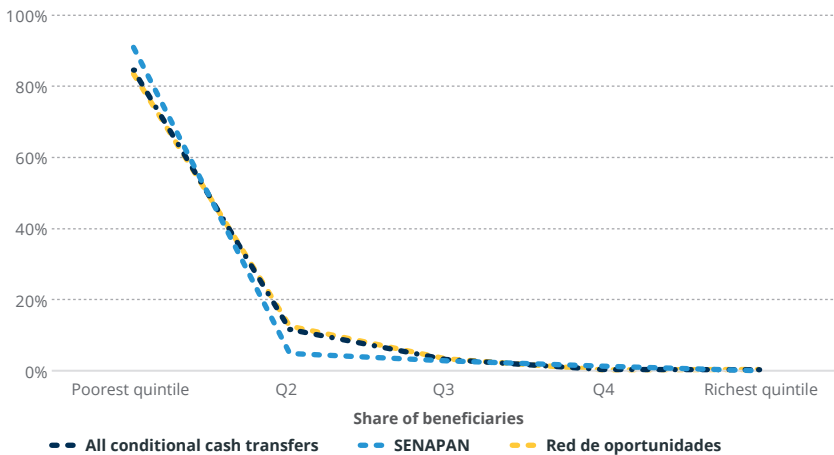
Source: ASPIRE database, www.worldbank.org/aspire.

Note: Figure based on conditional cash transfers (CCT) programs captured in Panama's 2018 Encuesta de Mercado Laboral. Coverage is: (Number of individuals in a given group [for example, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Figure 2C.2   Coverage and Beneficiary Incidence of Unconditional Cash Transfer Programs in South Africa, as Captured in the Household Surveys in ASPIRE, by Quintile of Pretransfer Welfare

a. Coverage of unconditional cash transfer programs
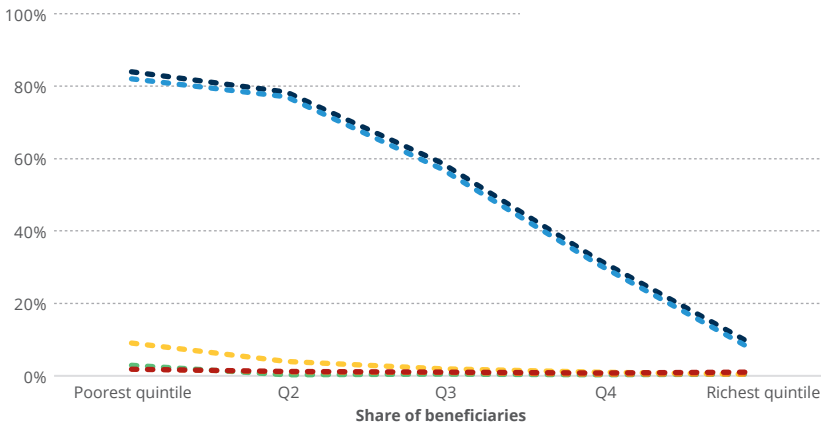
b. Incidence of unconditional cash transfer beneficiaries
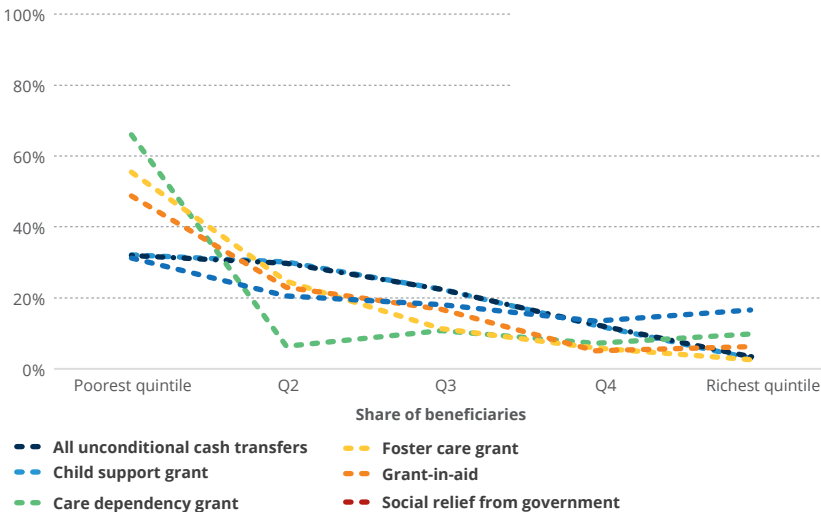
Source: ASPIRE database, www.worldbank.org/aspire.

Note: Figure based on unconditional cash transfers (UCT) programs captured in South Africa's 2014/15 Living Conditions Survey. Coverage is: (Number of individuals in a given group [for example, total population or poorest quintile] who live in a household where at least one member receives the transfer)/(Number of individuals in the group). Beneficiaries' incidence is: (Number of direct and indirect beneficiaries [people who live in a household where at least one member receives the transfer] in a given quintile)/(Total number of direct and indirect beneficiaries). The sum of percentages across quintiles per given instrument equals 100%. Quintiles are calculated using per capita pretransfer welfare (income or consumption). Q = quintile.

Annex 2D: Costs of Operating Social Registries

Annex Table 2D.1 Costs of Operating Social Registries

Country	Social registry information system	Budget cycle	Year	Static/ dynamic	Total cost of social registry (millions)	Number of programs using the social registry (Number)	Size of social registry Mil. fam./ people	Social registry average cost Per family	Monthly average benefit amount of a key program Average	Program	Total annual expenditure in social assistance programs (ASPIRE) (millions)	Expenditure year
Brazil	Cadastro Único	Annual	2020	Dynamic	US$188	20+	F: 28 P: 75	US$6.7	US$46.4	Bolsa Família in 2019	US$6,681.76	2018
	Registro Social de Hogares	Annual	2017	Dynamic	US$30	20+	F: 4.9 P: 12.9	US$6.1	US$18.25	Subsidio Único Familiar (SUF) in 2020	US$813.44	2017
Colombia	SISBEN IV	3-year cycle	2017–20	Static	US$21.9	20+	F: 8.5 P: 39	US$2.58	US$51	Familia en Accion in 2019	US$573.93	2018
	NEW Social Registry[a]	5-year cycle	2021–26	Dynamic	US$52.0	20+	F: 13 (target) P: 45.9 (target)	US$4 (for target)				
Republic of Congo	Lisungi Registry Social	3-year cycle	2017–19	Static	US$2.5	3+	F: 0.12 P: 0.6	US$20	US$25	Lisungi Cash Transfer in 2018	US$12.00	2019
Jordan	Unified National Register	Annual	2020	Dynamic	US$2.5	9	F: 1 P: 7	US$2.5	US$63 -200	National Aid Fund (NAF) in 2020	N/A	N/A
Malawi	UBR (phase 2)	N/A	2018	Static	US$1.27	5	F: 0.750 P: 3.5	US$1.7	US$8	Social Cash Transfer Programme (SCTP) in 2018	N/A	N/A
Mali	RSU (only counting those from Jigisemejiri program that represents 24% of RSU)	4-year cycle	2017–20	Static	US$1.68	8	F: 0.092 P: 0.525	US$18.3	US$20	Jigisemejiri Cash transfer program in 2019	N/A	N/A

Annex Table 2D.1 *(continued)*

Country	Social registry information system	Budget cycle	Year	Static/dynamic	Total cost of social registry (millions)	Number of programs using the social registry Number	Size of social registry Mil. fam./people	Social registry average cost Per family	Monthly average benefit amount of a key program Average	Program	Total annual expenditure in social assistance programs (ASPIRE) (millions)	Expenditure year
Pakistan	National Socio-Economic Registry (NSER)[b]	5-year cycle	2016–21	Static	US$56	70+	F: 25.5[c] P: 167	US$2.19	US$14	Kafaalat[d]	US$1,036.62	2016
Philippines	Listahanan	3-year cycle	2019–21	Static	US$54	20+	F: 16 P: 75	US$3.4	US$11 – 32	Pantawid program in 2019	US$1,080.47	2016
Senegal	Registre Social	4-year cycle	2019/20-2023/24	Static	US$3	3	F: 0.56 P: 4.8	US$5.36	US$16	Programme National des Bourses de Sécurité Familiale (PNBSF) in 2019	N/A	N/A
Tunisia	AMEN	4-year cycle	2016–20 (still on-going)	Static	US$ 3-4[11]	3+ (TBC)	F: 1 P:3.6 (target)	US$3-4	US$60	Permanent cash transfer program (PNAFN) in 2019	US$212.69	2015

Source: Original data collection based on informal survey of World Bank staff leading projects supporting the development of these social registries in consultation with government officials.

Note: AMEN = safety (in Arabic); N/A= not available.

a. In terms of municipalities support to SISBEN, there are more than 2,100 people who guarantee the day-to-day operation and implementation of the SISBEN. The costs associated with these staffs are under the responsibility of municipalities, and they add to about col$29 billion per year (US$8.8 million per year using 2020 conversion factor).
b. NSER administrators are already testing new functionalities to move toward a dynamic system, and some of future functionalities as part of the response of the country to COVID-19 pandemic.
c. NSER unit is household. On average, NSER found 1.3 families per household.
d. Kafaalat is the main program in Pakistan as part of the Ehsaas, which is the broader Social Protection and Poverty Alleviation Program launched in March 2019. Kafaalat was previously called Unconditional Cash Transfer under Benazir Income Support Program–BISP; BISP is now one of the programs under Ehsaas)

Notes

1. The estimates are simple in the sense that they do not take into account behavioral responses to the programs.
2. http://www.worldbank.org/aspire/.
3. The social assistance literature has generated a huge range of impact evaluations on a long vector of outcomes related to health, mental health, education, work, and livelihoods that programs can have through increasing income via transfers or the various and often associated elements of messaging and information, behavioral nudges, conditions for service use, job search, or work that can be implicit or explicit design elements of social assistance programs. This vector of outcomes is of paramount interest in social policy but beyond the subject of this book. It is the subject of many other papers and reviews; thus, a synopsis of that literature is not provided.
4. ADePT Social Protection is a free software platform developed by the World Bank to automate the generation of an array of indicators to assess the performance of social protection programs.
5. Less any transaction costs incurred from having applied to a program or psychic costs from having been excluded from the program.
6. Yemtsov et al. (2018).
7. http://www.worldbank.org/aspire/.
8. The change in the poverty gap gives more weight to a transfer to a very poor person than to one just below the poverty line. In that sense, it moves toward the ideal of continuous welfare weighting. The change in the poverty gap is commonly used and intuitive to understand. It falls short of ideal as it truncates consideration at the poverty line/welfare threshold. As described in chapter 7, alternative measures, such as distributional characteristics, would be preferable, but they are not regularly captured in ASPIRE's cross-country work, so they are not used here.
9. The World Bank's *State of Safety Nets 2018* report contains data on coverage and incidence for 96 countries since 2008 (World Bank 2018). Since it was published, more recent data have become available for 54 countries already in the database and two countries were added.
10. The countries were classified by income group using their standing at the time the analysis was done in May 2020.
11. The household surveys only capture national programs of significant size, and the subset of programs that target individuals, families, or households. Some of the social assistance programs captured in the ASPIRE administrative database include support for institutionalized social services, for example, for children deprived of parental care, persons with disabilities, or the elderly; these beneficiaries are not covered in the sample frame of a household survey.
12. For countries that happen to observe in their household questionnaire only a single program in a category, it is possible to look at the lines as program specific. The study team examined the documentation for every country and program category to see whether meaningful results could be produced by looking at that subset. It seemed that the programs and countries that were observed

were not representative of the larger world experience of social assistance. Single programs are much more likely to be observed in the categories of conditional cash transfer and school feeding programs.

13. ASPIRE focuses on data from traditional International Bank for Reconstruction and Development and International Development Association client countries, so its coverage of high-income countries is not the full set for the world (for example, the earliest industrializers, Europe, North America, Japan, Australia, and so forth). Thus, the higher income sometime-borrowers from the World Bank for which ASPIRE has reported survey data since 2014 are Chile, Croatia, Mauritius, Panama, Poland, Romania, and Uruguay.

14. From here, the category "other social assistance" is dropped because it covers such a heterogenous mix of programs and designs that it is difficult to draw general conclusions.

15. Beneficiary incidence is the proportion of beneficiaries in each quintile of the population's welfare distribution. Benefit incidence is the transfer amount received by each quintile as a percentage of total transfers received by the population.

16. Ley Nº 15.728h, Republic of Uruguay.

17. The story of Mongolia's Child Money Program has evolved over time. After 2016, the program drew on general revenues for finance and became targeted with varying degrees of coverage. However, by 2019 (and currently), it is back to being universal (UNICEF 2020).

18. This is the poverty rate using the international standard of $3.20/day/person rate for lower-middle-income countries; national poverty lines use a different standard.

19. Two UCT programs are observed in Montenegro's 2014 Household Budget Survey: the guaranteed minimum income program and child allowances. However, the variable for child allowances only reports six observations; therefore, this section considers the UCT category as mainly representing the guaranteed minimum income program.

20. This subsection only uses observations for which there is clear valuation of the benefit levels. This restricts the analysis to the programs with benefits in cash—UCT, conditional cash transfers, social pensions, and public works.

21. https://socialprotection.org/discover/programmes/asignaciones-familiares -plan-equidad-family-allowances-equity-plan.

22. Two conditional cash transfer programs are observed in Panama's 2018 Encuesta de Mercado Laboral: Red de Oportunidades, which covers 5.7 percent of the total population, and SENAPAN, a smaller program covering less than 1 percent of the population. Panama also has two additional programs that target poor individuals: Programa Ángel Guardián (for people with disabilities) and B/.120 a los 65 (for adults 65 years and older who do not receive a contributory pension). Although Panama's government considers that these programs are conditional cash transfers due to the inclusion of conditions, ASPIRE classifies them as social pensions since their conditions are not related to investments in human capital, such as school attendance, immunizations, health check-ups, and so forth. See annex 2C.

23. Costs can also be recurrent or one-time capital investments. In the life of a social assistance program, there is an inverted U shape in the evolution of the costs of administration. The costs can be uneven—high startup costs initially spread over initially small coverage and then decreasing as share over time as the numerator decreases and the denominator increases; the periodic survey sweep model gives uneven costs.

24. See Tesliuc et al. (2014) for a good discussion of this issue and sample questionnaires for the special data collection that is needed to obtain the best estimates of costs.

25. In some areas of Burkina Faso and in rural Chad, when program administrators had collected data for all households in selected areas in the program registry to then apply the targeting method, the shares of administrative costs were 1.5 and 3.9 percent, respectively. These costs were much lower, at 0.4 percent in other areas of Burkina Faso and 1.4 percent in urban Chad, when household data were collected only from those who requested to participate (self-selection) in the program. In Niger, the authors compare two approaches: full data collection for all households in selected areas, and partial data collection after the community defines the pre-list of potential beneficiaries. The shares of administrative costs in total program costs were similar, 5.5 percent for the former and 4.3 percent for the latter.

26. These figures were obtained through an informal survey of World Bank staff leading projects supporting the development of these social registries in consultation with government officials. See annex 2D.

27. Devereaux et al. (2017).

28. In Mali, the unit cost of a questionnaire for the Registre Social Unifie (www .rsu.gouv.ml) collected by the Jigisemejiri program has dropped from US$20 in 2017–20 to US$10 today.

29. Turkey's population was estimated at 78 million in 2015.

30. Turkey's Integrated Social Assistance System Report, http://documents1 .worldbank.org/curated/en/515231530005107572/pdf/Turkey-SA-summary. pdf.

31. https://veja.abril.com.br/blog/radar/governo-tirou-13-milhao-de-beneficiarios -irregulares-do-bolsa-familia/.

References

Acosta, Pablo A., Phillippe Leite, and Jamele Rigolini. 2011. "Should Cash Transfers Be Confined to the Poor? Implications for Poverty and Inequality in Latin America." Policy Research Working Paper 5875, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/3642.

Adato, Michelle. 2000. "The Impact of *PROGRESA* on Community Social Relationships." Final Report. International Food Policy Research Institute, Washington, DC. https://www.ifpri.org/publication/impact-progresa-community -social-relationships.

Attah, Ramlatu, Valentina Barca, Andrew Kardan, Ian MacAuslan, Fred Merttens, and Luca Pellerano. 2016. "Can Social Protection Affect Psychosocial

Wellbeing and Why Does This Matter? Lessons from Cash Transfers in Sub-Saharan Africa." *Journal of Development Studies* 52 (8): 1115–31. https://www.tandfonline.com/doi/full/10.1080/00220388.2015.1134777.

Baird, Sarah, David McKenzie, and Berk Özler. 2018. "The Effects of Cash Transfers on Adult Labour Market Outcomes." *IZA Journal of Development and Migration* 8 (22). https://izajodm.springeropen.com/articles/10.1186/s40176-018-0131-9.

Banerjee, Abhijit V., Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken. 2017. "Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programmes." *World Bank Research Observer* 32 (2): 155–84. http://hdl.handle.net/10986/31221.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano. 2016. "Cash Transfers: What Does the Evidence Say? A Rigorous Review of Programme Impact and of the Role of Design and Implementation Features." Overseas Development Institute, London. https://odi.org/en/publications/cash-transfers-what-does-the-evidence-say-a-rigorous-review-of-impacts-and-the-role-of-design-and-implementation-features.

Baumberg, Ben. 2016. "The Stigma of Claiming Benefits: A Quantitative Study." *Journal of Social Policy* 45 (2): 181–99. https://www.cambridge.org/core/journals/journal-of-social-policy/article/stigma-of-claiming-benefits-a-quantitative-study/AF30092AE7D5B7C659798228B219F02C.

Beegle, Kathleen, Aline Coudouel, and Emma Monsalve. 2018. *Realizing the Full Potential of Social Safety Nets in Africa*. Washington, DC: World Bank. http://hdl.handle.net/10986/29789.

Besley, Timothy, and Ravi Kanbur. 1990. "The Principles of Targeting." DEC Working Paper 385, World Bank, Washington, DC.

Brown, Caitlin Susan, Martin Ravallion, and Dominique van de Walle. 2016. "A Poor Means Test? Econometric Targeting in Africa." Policy Research Working Paper 7915, World Bank, Washington, DC. http://hdl.handle.net/10986/25814.

Caldes, Natalia, David Coady, and John Maluccio. 2006. "The Cost of Poverty Alleviation Transfer Programs: A Comparative Analysis of Three Programs in Latin America." *World Development* 34 (5): 818–37. https://EconPapers.repec.org/RePEc:eee:wdevel:v:34:y:2006:i:5:p:818-837.

Castaneda, Tarcisio, and Kathy Lindert. 2005. "Designing and Implementing Household Targeting Systems: Lessons from Latin America and the United States." Social Protection Discussion Paper Series No. 0526, World Bank, Washington, DC. https://imagebank2.worldbank.org/search/6031977.

CGD (Center for Global Development). 2015. "South Africa's Child Support Grant." CGD, Washington, DC. http://millionssaved.cgdev.org/case-studies/south-africas-child-support-grant.

Cornia, Giovanni, Richard Jolly, and Frances Stewart. 1987. *Adjustment with a Human Face*. Clarendon Press.

Daigneault, Pierre-Marc, Steve Jacob, and Maximilien Tereraho. 2012. "Understanding and Improving the Take-up of Public Programs: Lessons Learned from the Canadian and International Experience in Human Services." *International Journal of Business and Social Science* I (3): 39–50. http://ijbssnet.com/view.php?u=http://ijbssnet.com/journals/Vol_3_No_1_January_2012/5.pdf.

Delany, Aislinn, and Selwyn Jehoma. 2016. "Implementation of Social Grants: Improving Delivery and Increasing Access." In *South African Child Gauge 2016*, edited by Aislinn Delany, Selwyn Jehoma, and Lori Lake, 60–67. Cape Town: Children's Institute, University of Cape Town. http://www.ci.uct.ac.za/sites/default/files/image_tool/images/367/publication/2016/implementation-of-social-grants.pdf.

Della Guardia, Anne, Milli Lake, and Pascale Schnitzer. Forthcoming. "Poverty Targeting When Everyone Is Poor: The Unintended Consequences of Selective Inclusion in Cash Transfer Programs." World Bank, Washington, DC.

Devereux, Stephen, Edoardo Masset, Rachel Sabates-Wheeler, Michael Samson, Althea-Maria Rivas, and Dolf te Lintelo. 2017. "The Targeting Effectiveness of Social Transfers." *Journal of Development Effectiveness* 9 (2): 162–211. http://dx.doi.org/10.1080/19439342.2017.1305981.

DSD, SASSA, and UNICEF (Department of Social Development of South Africa, South African Social Security, and United Nations Children's Fund). 2012. *The South African Child Support Grant Impact Assessment: Evidence from a Survey of Children, Adolescents and Their Households*. Pretoria: UNICEF South Africa. https://www.unicef.org/southafrica/media/1116/file/ZAF-South-African-child-support-grant-impact-assessment-2012.pdf.

Gotcheva, Boryana, Aylin Isik-Dikmelek, Matteo Morgandi, Victoria Strokova, Tomas Damerau, and Vojin Golubovic. 2013. "Activation and Smart Safety Nets in Montenegro: Constraints in Beneficiary Profile, Benefit Design, and Institutional Capacity." World Bank, Washington, DC.

Granlund, Stefan, and Tessa Hochfeld. 2020. "'That Child Support Grant Gives Me Powers': Exploring Social and Relational Aspects of Cash Transfers in South Africa in Times of Livelihood Change." *Journal of Development Studies* 56 (6): 1230–44. https://doi.org/10.1080/00220388.2019.1650170.

Grosh, Margaret, Carlo del Ninno, Emil Tesliuc, and Azedine Ouerghi. 2008. *For Protection and Promotion: The Design and Implementation of Effective Safety Nets*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/6582.

Gubrium, Erika, and Sony Pellissery. 2016. "Antipoverty Measures: The Potential for Shaming and Dignity Building through Delivery Interactions." *International Journal of Social Quality* 6 (2):1–17. https://www.semanticscholar.org/paper/Antipoverty-Measures%3A-The-Potential-for-Shaming-and-Gubrium-Pellissery/cd80a785964977f794bb8cb7b03ac5a2b8c8c6d1.

Hanna, Reema, and Ben Olken. 2018. "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries." *Journal of Economic Perspectives* 32 (4): 201–26. https://www.aeaweb.org/articles?id=10.1257/jep.32.4.201.

ILO (International Labour Organization). 2014. *Namibia Social Protection Floor Assessment*. Geneva: ILO. https://www.social-protection.org/gimi/gess/ShowRessource.action?ressource.ressourceId=47837.

ILO (International Labour Organization) 2016. "Mongolia Child Money Programme." Social Protection Department, ILO, Geneva. https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---ilo-beijing/documents/publication/wcms_534930.pdf.

Jaramillo, Miguel, and Alan Sánchez. 2011. *Impacto del Programa Juntos sobre Nutrición Temprana.* Documento de Investigación 6, GRADE, Lima. http://www.grade.org.pe/en/publicaciones/1038-impacto-del-programa-juntos-sobre-nutricion-temprana/.

Jones, Nicola, Rosana Vargas, and Eliana Villar. 2007. "Conditional Cash Transfers in Peru: Tackling the Multi-Dimensionality of Poverty and Vulnerability." In *Social Protections Initiatives for Families, Women and Children: An Analysis of Recent Experiences*, edited by Alberto Minujin. New York: The New School with support from UNICEF. https://cdn.odi.org/media/documents/1678.pdf.

Kandpal, Eeshani, Pascale Schnitzer, and Modeste Daye. 2020. "The Effects of Cash Transfers on Women's Productive Activities." World Bank, Washington, DC.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach. 2018. "Does Poverty Lower Productivity?" NBER Working Paper 28338, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w28338.

Kidd, Stephen, and Diloá Athias. 2020. "Hit and Miss: An Assessment of Targeting Effectiveness in Social Protection." Working Paper, Development Pathways, London. https://www.developmentpathways.co.uk/wp-content/uploads/2019/03/Hit-and-miss-long-report-.pdf.

Kidd, Stephen, Diloá Athias, and Idil Mohamud. 2021. "Social Registries: A Short History of Abject Failure." Working Paper, Development Pathways, London. https://www.developmentpathways.co.uk/wp-content/uploads/2021/06/Social-registries-a-short-history-of-abject-failure-June-1.pdf.

Knox-Vydmanov, Charles. 2011. "The Price of Income Security in Older Age: Cost of a Universal Pension in 50 Low- and Middle-Income Countries." Discussion Paper, HelpAge International, London. https://www.helpage.org/silo/files/the-price-of-income-security-in-older-age.pdf.

Kosec, Katrina, and Cecilia Hyunjung Mo. 2017. "Aspirations and the Role of Social Protection: Evidence from a Natural Disaster in Rural Pakistan." *World Development* 97: 49–66. https://www.sciencedirect.com/science/article/pii/S0305750X17301092.

Leite, Phillippe, Tina George, Changqing Sun, Theresa Jones, and Kathy Lindert. 2017. "Social Registries for Social Assistance and Beyond: A Guidance Note and Assessment Tool." Social Protection and Labor Discussion Paper No. 1704, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/28284.

Lindert, Kathy, Colin Andrews, Chipo Msowoya, Boban Varghese Paul, Elijah Chirwa, and Anita Mittal. 2018. "Rapid Social Registry Assessment: Malawi's Unified Beneficiary Registry." Social Protection and Jobs Discussion Paper No. 1803, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/31012.

Lindert, Kathy, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Nishikawa Chávez. 2020. *Sourcebook on the Foundations of Social Protection Delivery Systems*. Washington, DC: World Bank. http://hdl.handle.net/10986/34044.

Moffitt, Robert. 2015. "Economics of Means-Tested Transfer Programs: Introduction." Working Paper 21751, National Bureau of Economic Research, Cambridge, MA. https://www.nber.org/papers/w21751.

Oliveira, Victor, Laura Tiehen, Mark Prell, and David Smallwood. 2018. "Evolution and Implementation of the Supplemental Nutrition Assistance Program in the United States." In *The 1.5 Billion People Question: Food, Vouchers, or Cash Transfers?* edited by Harold Alderman, Ugo Gentilini, and Ruslan Yemtsov. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/27907.

OPM (Oxford Policy Management). 2014. "Qualitative Research and Analyses of the Economic Impacts of Cash Transfer Programmes in Sub-Saharan Africa: Lesotho Country Case Study Report." Protection to Production, Food and Agriculture Organization, Rome.

Packard, Truman, Ugo Gentilini, Margaret Grosh, Philip O'Keefe, Robert Palacios, David Robalino, and Indhira Santos. 2019. *Protecting All: Risk Sharing for a Diverse and Diversifying World of Work.* Human Development Perspectives. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/32353.

Perova, Elizaveta, and Renos Vakis. 2009. "Evaluating the Juntos Program in Peru: Evidence from Non-Experimental Estimates." World Bank, Washington, DC.

Perova, Elizaveta, and Renos Vakis. 2012. "5 Years in Juntos: New Evidence on the Program's Short and Long-Term Impacts." *Economía* (Departamento de Economía, Pontificia Universidad Católica del Perú) 35 (69). https://www.researchgate.net/publication/254441506_5_Years_in_Juntos_New_Evidence_on_the_Program's_Short_and_Long-Term_Impacts.

Plagerson, Sophie, and Marianne Ulriksen. 2015. "Cash Transfer Programmes, Poverty Reduction and Empowerment of Women in South Africa." Conditions of Work and Equality Department, Gender, Equity and Diversity Branch, International Labour Office, Geneva. https://www.researchgate.net/publication/285322517_Cash_transfer_programmes_poverty_reduction_and_empowerment_of_women_in_South_Africa.

Republic of Ghana. 2018. "Ghana LEAP 1000 Programme: Endline Evaluation Report." Ministry of Gender, Children and Social Protection, Republic of Ghana, Accra. https://www.unicef-irc.org/files/documents/d-4108-LEAP-1000-Report.pdf.

Republic of Montenegro. 2013. "Law on Social and Child Welfare." Ministry of Labor and Social Welfare, Republic of Montenegro. http://ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=95155&p_count=96232&p_classification=04&p_classcount=2269.

Roelen, Keetie. 2017. "Shame, Poverty and Social Protection." IDS Working Paper No. 489, CSP Working Paper No. 015, Centre for Social Protection, Institute of Development Studies, Brighton, UK. https://www.ids.ac.uk/publications/shame-poverty-and-social-protection/.

Roelen, Keetie. 2020. "Receiving Social Assistance in Low- and Middle-Income Countries: Negating Shame or Producing Stigma?" *Journal of Social Policy* 39 (4): 705–23. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14772.

Rosas, Nina, Samantha Zaldivar, and Mariana Pinzon-Caicedo. 2016. "Evaluating Tanzania's Productive Social Safety Net: Targeting Performance, Beneficiary Profile, and Other Baseline Findings." World Bank, Washington, DC. 10.13140/RG.2.2.12257.97120.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan. 2016. "The Psychological Lives of the Poor." *American Economic Review* 106 (5): 435–40. https://www.aeaweb.org/articles?id=10.1257/aer.p20161101.

Schnitzer, Pascale, and Quentin Stoeffler. 2020. "Targeting in Practice: Evidence from Nine Programs in Africa." Policy Research Working Paper 9816, World Bank, Washington, DC. https://quentinstoeffler.weebly.com/uploads/4/0/2/6 /40265181/sahel_targeting_spj_bbl_11052020.pdf.

Tesliuc, Emil, Lucian Pop, Margaret Grosh, and Ruslan Yemtsov. 2014. *Income Support for the Poorest: A Review of Experience in Eastern Europe and Central Asia*. Directions in Development: Human Development. Washington, DC: World Bank. http://hdl.handle.net/10986/18886.

Thinkthrough. 2021. "Analytical Deep Dive on Social Security Administration Registration in Nepal." Draft Report, Thinkthrough Consulting, New Delhi, India.

UNICEF (United Nations Children's Fund). 2020. *Universal Child Benefit Case Studies: The Experience of Mongolia*. New York: UNICEF. https://www.unicef.org /media/70471/file/MNG-case-study-2020.pdf.

World Bank. 2018. *The State of Social Safety Nets 2018*. Washington, DC: World Bank. http://hdl.handle.net/10986/29115.

World Bank. 2019. "Public Expenditure Review of Social Protection in Jamaica: Principal Programs, Services, and Instruments." World Bank, Washington, DC.

Wright, Gemma, David Neves, Phakama Ntshongwana, and Michael Noble. 2015. "Social Assistance and Dignity: South African Women's Experiences of the Child Support Grant." *Development Southern Africa* 32 (4): 443–57. https://www .tandfonline.com/doi/full/10.1080/0376835X.2015.1039711.

Yang, Lichao, Robert Walker, Jiexuan Chen, and Rui Zhang. 2019. "Universal Child Benefits and Dignity and Shame." UNICEF Child Benefits Research Project, United Nations Children's Fund, New York. https://socialprotection.org/sites /default/files/publications_files/Universal%20Child%20Benefits%20and%20 Dignity%20and%20Shame.pdf.

Yemtsov, Ruslan, Maddalena Honorati, Brooks Evans, Zurab Sajaia, and Michael Lokshin. 2018. *Measuring the Effectiveness of Social Protection: Concepts and Applications. Streamlined Analysis with ADePT Software*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/29802.

3

Moving from General Abstraction toward Implementable Concepts in Stability and in Crisis

Margaret Grosh and Matthew Wai-Poi

Differentiating eligibility or benefits along some continuum of well-being is intuitive as a general idea, but in moving from an abstract vision to more specific concepts to implementable definitions and procedures, several topics require careful thought. This chapter provides a synopsis of the key issues. It presents the details in simple terms, distilling lessons from the vast body of work on poverty economics and survey science and connecting them to social assistance practice. The chapter's high level of generality is meant to inform the general choices and rules that a social protection delivery system would aim to implement. Chapter 4 takes up greater details on the delivery system, chapter 5 discusses the implications for the choice of targeting method, and chapter 6 explains the how-to of data and inference.

   The chapter is organized in two parts. The first part clarifies the concepts and presents a range of choices for action in stable environments. It is written around a series of questions important in designing and implementing social protection programs. For example, social protection for whom? Whose welfare is measured—an individual's or a family's? What defines a family? What is the system trying to measure? Where does it draw the line(s) between eligibility and not?

149

The second part introduces dynamics and shocks. It notes that welfare is dynamic, with many people being somewhat better or worse off from time to time. Even idiosyncratic changes add a layer of challenges for social assistance programs. Large covariate shocks, such as natural disasters and economic crises, step those up considerably. The chapter reviews ideas and highlights fruitful approaches with recent experiences.

Part I: Even in Times of Stability, Welfare Measurement for Eligibility Determination Is Complex

Since social assistance is often motivated by the need to address poverty and inequality, the literatures are intertwined. The analytic literature shapes the agenda on the issues that must be decided (and sometimes reexamined) as programs move to implementation.

For Whom Are We Trying to Measure Welfare?

Poverty assessments take the "household" as the basic unit of data collection and analysis—people living together in a dwelling and sharing basic expenses or cooking together,[1] with welfare (usually measured by consumption or income) usually represented in per capita terms, sometimes with adjustments for adult equivalence and/or economies of scale in household size.[2] Similar considerations of adjustments to eligibility thresholds and benefit levels may be made in social assistance programs.

In social assistance, the "assistance unit" is not always defined in the way that surveys define the "household."[3] Sometimes the assistance unit is an individual—a child for a child allowance, an elderly person for a social pension, or a person living with disability for a disability allowance. However, often the assistance unit is the nuclear family—parent(s) and children—although nuclear families can live in larger compound households, especially traditional multigenerational families that incorporate grandparent(s) with at least one adult child with a spouse and children. In the 89 developing countries with data available in the Global Monitoring Database, nuclear families (a couple and children) comprise only 31 percent of households, and 41 percent of those with welfare below the $1.90/day/person poverty line (Munoz Boudet et al. 2018). Thus, social assistance policy must consider many more complex family structures (see figure 3.1).

The definition of the assistance unit is important for eligibility decisions, as examination of a basic multigenerational household illustrates. Consider a household with an elderly couple, an adult son and his wife and children, and a young adult daughter. The adult son is formally employed and has a middle-class income. The dwelling is commensurately nice and registered

Figure 3.1 Distribution of Poor (by the $1.90 Metric) Households, by Demographic Household Typology
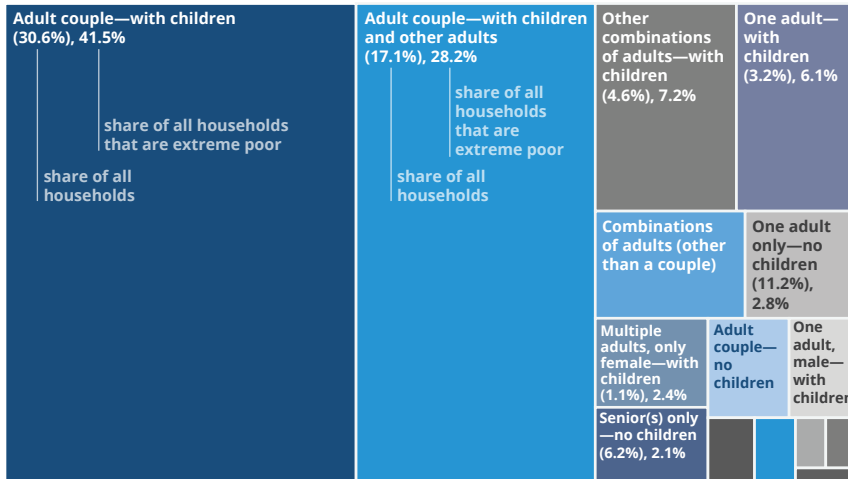
Source: Munoz Boudet et al. 2018.

in the name of the adult son (and maybe his wife). The adult daughter earns an occasional income, maybe babysitting or as a temp worker. If the members of this household are all viewed as one assistance unit, then all income and assets will be pooled, and any economies of age and scale assigned. The son's good steady income or caliber of housing might find the household nonneedy if viewed as a whole. But if viewing the household members as three units (elderly couple, nuclear family, and single individual), the calculus could change. The young adult daughter could be viewed as having small, irregular income and no housing assets. She might be eligible for income support or job training or placement. The elderly couple might be eligible for a means-tested social pension. The bureaucratic/administrative concept of an assistance unit may be related to the purposes of a program or to the nationally pertinent vision of what constitutes a "normal" or "good" life in terms of family structure and independence.

Types of Households

Although who shares a dwelling has a basis in culture and family bonds, it also responds to economics. That is, nuclear families or individuals move in and out in response to economic opportunities or pressures, and this has implications for eligibility determination processes. In addition to multigenerational households and whether to consider them as one or more assistance units, social assistance policy must consider various other fairly

common situations. Defining optimal rules involves difficult choices with possible trade-offs, and so there is no single blanket recommendation that applies across all circumstances or that all people would agree on in a given circumstance. Understanding how social assistance and household formation interact may be important for equity, for the goals of the program, and to understand possible unintended consequences.

Nonmarital unions. Formal marriages provide a "base case" for policy as they are clearly defined, formally recorded, and pair norms of affection and economic support. But real life is often less clear-cut, and in many countries, formal marriage is not ubiquitous or it is on the decline.[4] Thus, welfare policy must account for varied living situations and how to handle eligibility determination for each. For example, should the noncustodial parent's child support payments count as part of the family's income in the social assistance unit in which the child resides? Should such payments be discounted from the income of the paying noncustodial parent? How can the social assistance administrator track child support payments? How often/by how much would social assistance benefits be adjusted if child support were not paid regularly? Should the income of a resident (new) partner of the custodial parent be counted? And if so, how is residency or partnership determined? The wording of these questions shows that they pertain in more force for means-tested programs trying to fine-tune measures of income than for proxy means-tested programs that rely on markers of housing characteristics and assets. The questions show the possible imprecisions in thinking about unmarried mothers or even widows as necessarily being reliant on only their own single incomes as some may have informal partners who contribute to the expenses of a shared household.

The issue of how to define social assistance policy to meet various goals—to treat couples of similar economic means similarly irrespective of the formality of their union, to encourage/not discourage marriage, and to avoid intrusive inspections or unclear rules—has been a significant and much debated issue. This has been especially true of US welfare policy for several decades, but the issue pertains in other contexts as well (see, for example, Moffitt, Phelan, and Winkler [2020] and Wilcox, Price, and Rachidi [2016]). Various European countries have moved further than the United States has toward reducing "marriage penalties" by considering the income of nonmarital partners, but countries still struggle with finding definitions that are reasonably clear and observable (Besharov and Gilbert 2015).

Polygamous households. There is no precise tally of the practice of definition of social assistance units for polygamous households, but the impression is that it is not uncommon but not ubiquitous in social assistance programs to define the assistance unit as the combined household of a husband, all wives, their children, and any additional members, especially where they live together in a compound. To treat the household as one

assistance unit implies a high degree of sharing of resources, although some research suggests that resources are not fully pooled among husbands and co-wives,[5] with the implication that treating each wife and her children as an assistance unit may be desirable, especially for programs aiming at direct improvements of children's human capital outcomes. However, this option may be more difficult in a data collection sense.[6]

Informal foster children.[7] Children may not always live with their biological parents. For example, older children may live with other relatives, at least for defined time periods. Those relatives may be able to provide better material care, such as more food or closer access to schools, or even out caregiving or household labor while the children are still within the embrace of (extended) family (see, for example, Akresh 2009). Such traditions were strong especially in Sub-Saharan Africa[8] and became an important means of coping when the HIV/AIDS epidemic hit the continent so hard, with not only children who lost both parents but even many of those who lost one residing with family other than the surviving parent (see, for example, Ardington and Liebrandt [2010] and Penglase [2021]). Fluidity in family structure makes it all the more important to understand how to target social policy—for example, to orphans per se, to all children in families that host orphans, or to children in poor families more generally, especially as the overlap varies substantially among countries (Ainsworth and Filmer 2002). It is possible that COVID-19 will bring a new increment to such arrangements and across a wider set of countries due to families facing the loss of parental or grandparental caregivers. Hillis et al. (2021) and Unwin et al. (2022) estimate that globally, in the first 20 months of the pandemic (up to October 31, 2021), 5.0 million COVID-19 deaths had occurred, and 5.2 million children had lost a parent or caregiver due to COVID-19–associated deaths; 3 of every 4 children affected by orphanhood lost their fathers; 2 of every 3 children whose parents died were adolescents. Affected countries cover the globe, with highest rates of paternal orphanhood in India, Mexico, Peru, and South Africa.

"Guest" households. At the time of this writing, COVID-19–related eviction protections are expiring in the United States. It is expected that with many workers having lost their jobs or suffered huge reductions in earnings, some of the families they support will lose the housing they could afford before the crisis. Some people may be able to find lower cost housing and continue to live independently, and some (hopefully very few) may end up homeless. A goodly share will probably move in with others—family, friends, or roommates.[9] This shock absorption has worked around the world in response to many previous waves of economic shocks, natural disasters, and migrations. (In Spanish, the term for the economically needy "guest" households lodging within a previously established one is *los allegados*, or those who have arrived.) The new arrivals' "identity" and

family finances will remain at least somewhat and maybe fully separate from those of the "host" family in which they land, at least initially and probably more markedly when the prior degree of autonomy was greater or social relationship weaker. The social assistance policy in some countries considers the guest and host families as qualifying as separate assistance units. Operationalizing that is fairly straightforward for means-tested programs, by counting the income of each member of the assistance unit. For programs that base eligibility on geography or consider the characteristics of the dwelling as proxies for welfare, it is more difficult to operationalize. Such programs normally measure the characteristics of the host family and do not reflect the independent welfare of the guest family. This situation may be tricky for programs that consider housing need—is the guest family well lodged or a priority for shelter or housing programs?

Combined employer/staff households. In some countries, the employment of live-in domestic staff is not uncommon. Socially, there is little question that the employers and staff belong to different assistance units, and, in general, families that are well-off enough to employ live-in domestic staff are not the population of interest for many social assistance programs. But their staff may be, especially the children of the staff. Thus, issues that are similar to those of guest households arise. Will the targeting method be able to differentiate the welfare of the staff household given that it may reside in a dwelling with amenities in a good neighborhood?

The housing unit. Occasionally, it is a feature of the dwelling such as the electricity or water meter that de facto defines the assistance unit, for example, programs that provide reduced tariffs (or even outright transfers) to those whose meters record low usage. Here the meter is what is observable, not how many people use it or how closely they are related. The use on the meter is an easily observed metric of welfare, but it becomes less precise to the degree that meters are shared with multiple households; poorer households, which are likely to be larger and thus use more water or electricity, or to be part of pirate connection schemes; or households that have no utilities at all (see, for example, Komives et al. [2005]; Ruggeri Laderchi [2014]).[10]

Individuals

Several of the common social assistance programs are named with a focus on individuals rather than families or households—child allowances, social pensions, and disability and veterans' benefits. In practice, some programs with such names rely only on categorical targeting and so really have only the individual as the unit of social assistance. However, many programs also have some sort of differentiation of eligibility or benefits by economic need as well. Thus, such programs require all the details for defining individuals' wider social assistance units and measuring their welfare.

There are several reasons why programs might be focused on individuals rather than families or households. It is important to be clear about which reasons carry what degree of weight in a given setting to think through the eligibility mechanism that may match. For some reasons, categorical targeting matches well, and for other reasons, it may not.

Many programs with the individual as the social assistance unit are related to recognition of special merit or vulnerability. Social merit can comprise groups such as war veterans, families of war victims or atrocities, "hero" mothers, and occasionally even artists. Such programs are not uncommon, but usually they are fairly small and, if related only to merit, not a branch of social policy on which this book can shed much light. Programs for children and people living with disability often cite a strong basis in rights and documents, such as the Convention on the Rights of the Child (1989), the Convention on the Rights of Persons with Disabilities (2006), or the movement on the rights of the elderly.[11] The vulnerabilities of these groups are well understood (box 3.1). The consensus around these rights is strong, as testified by the number of signatory nations. Since the rights established for these groups are congruent with the economic and social rights established for wider populations, the dilemmas about progressive realization are similar to those discussed in chapter 1.

Often programs that nominally focus on individuals as the social assistance unit are grounded in a perceived correlation between the individual category and need. For example, veterans may receive preferential access to training or livelihood programs to smooth their transition to civilian life. Children and the elderly may not be expected to generate income and indeed may to varying degrees require care that reduces the earnings of caregivers. Thus, children and the elderly, as individuals or households with more than an average share of nonearners, may be poorer on average. Disability is often listed among categorical benefits, but it is much more complicated and in some ways unlike other categories, at least in the difficulty of determining who belongs in the category (see box 5.5, in chapter 5).

If correlation with need is the main reason for programs focused on individuals, then it is very important to consider in detail, quantitatively and in each pertinent setting, how strong such correlations are. This leads to several methodological questions, especially relating to economies of scale and equivalence, which can be particularly important in welfare rankings of families of different demographic compositions. While the theory of why both are sensible constructs to use is intuitive and broadly shared, consensus on a practical calibration of them remains elusive (Deaton 2016; Deaton and Zaidi 2002; Newhouse, Suarez-Becerra, and Evans 2016; Ravallion 2015; World Bank 2018a).

Varied Usages Given to the Words "Vulnerable" and "Vulnerability"

The word "vulnerable" is used in several different ways in poverty and social protection. It is important to be clear about which meaning applies, through clear word usage or context. Without such clarity (and too often in practice), discussions can become muddled. This is all the more so because although the concepts are clearly distinct, there may be co-occurrences of different sorts of vulnerability for the same individual or family.

- Children are biologically vulnerable in that their development is very sensitive to any deprivations (in nutrition, care, and physical and social environment). The elderly may have frail health. Children and some elderly individuals may also be vulnerable in the sense of limited agency. They depend on others to make decisions and act on their behalf.
- Individuals can be called socially vulnerable because they have a trait that is associated with exclusion, for example, being a member of a nondominant indigenous, ethnic, racial, or religious group; or living with a disability; or identifying as lesbian, gay, bisexual, transgender, intersex, or other in a context marked by discrimination with respect to education, jobs, services, and social interaction. Such discrimination is important in and of itself. Moreover, it may carry with it the likelihood of poverty among those with one or more of the traits associated with exclusion.
- Individuals are considered vulnerable to poverty in the sense of having an income that is close to the poverty line, having high variability of income, living in an area prone to shocks, or all three together.

In general, the finding that children or families with children are poorer is common, but the link between the elderly and poverty is less robust. For example, Guven and Leite (2016) examine elderly poverty in 12 countries in Sub-Saharan Africa. In most of the countries, the elderly are not poorer than other groups, a result that echoes similar findings by Kakwani and Subbarao (2005).[12] In looking at the contours of child poverty, Newhouse, Suarez-Becerra, and Evans (2016) show that, globally, among children younger than 18 years, 19.5 percent are estimated to live on less than $1.90 per day, as opposed to 9.5 percent of prime-age adults and 7.0 percent of those ages 60 and older, with similar patterns for higher poverty lines.[13]

The number of poor children is more than eight times the number of poor elderly. The authors test alternative scales of economies and equivalences and find that the fundamental result of higher child than adult poverty holds, irrespective of scaling. There is important variation across countries, and so policy makers will want to look at numbers specific to their countries (see chapter 5).

Most of the analyses that profile welfare and household composition are done on the basis of household survey data that observe welfare jointly for the whole household and do not look within the household. Work attempting to do so is more nascent, requiring much stronger modeling assumptions, more detailed household survey information that collects significant data at the individual rather than the household level, or both. The still somewhat limited available evidence suggests that there is hidden poverty among children and women, and to a lesser extent the elderly, in nonpoor families (Brown, Calvi, and Penglase 2018; World Bank 2018a). It also suggests that measures of wealth correlate surprisingly little with other important metrics of welfare, such as anthropometrics (Brown, Ravallion, and van de Walle 2019).

The social assistance unit has important ramifications for delivery systems. In general, countries have some sort of foundational identification (ID) registry that allows social programs to verify the unique identity of an individual, and the IDs belong to individuals irrespective of life changes. Thus, all the administrative functions of programs with the individual as the unit of assistance can refer to a natural and immutable ID number (despite the still salient challenges with respect to lack of coverage and centralized or digitized records). However, a family registry or a household registry system must be constructed and updated with marriages and divorces, births and deaths, and movements to age of majority/independence or out of the household. Since the definitions are more difficult (as discussed above) and changes in family or household composition are common, programs with the family or household as the unit of assistance have an added layer of complexity to handle in delivery systems. (See chapter 4 for more details.)

What Are We Trying to Measure?

Welfare—and lack of welfare, or poverty—is a multidimensional concept. This is now widely recognized and the World Bank (among others) recommends measuring outcomes on both monetary and nonmonetary dimensions. The nonmonetary measures that are often considered include housing, basic infrastructure (electricity, drinking water, sanitation, and cooking fuel), access to health and education services, and crime and security.

At the household level, two measures of multidimensional poverty have been estimated for most countries: the United Nations Development Programme and Oxford Poverty and Human Development Initiative's Global Multidimensional Poverty Index (Global MPI) and the World Bank's multidimensional poverty headcount rate.[14] The key difference between them is that the Global MPI includes only nonmonetary dimensions, while the World Bank measure includes both monetary and nonmonetary dimensions.

Food security is another nonmonetary measure of welfare that is commonly used. Food security measures whether people can buy food (is it available and affordable?) and benefit nutritionally from eating it (does it contain the right nutrients, can they prepare it properly, and can they metabolize it?). Food security can be defined and measured in different ways and is commonly used by the United Nations and other humanitarian actors to target aid. It is often correlated with monetary poverty but can differ for important reasons (see box 3.2).

Box 3.2

Food Security and Money-Metric Welfare

Food (in)security is a metric of welfare used to prioritize assistance during times of stability and times of shock. Given its widespread use, it is useful to understand its relationship to money-metric measures.

Food security covers several dimensions: (1) availability (is food available in a certain location?), (2) access (can a household access adequate food, given prices, incomes, and its access to formal and informal social assistance?), and (3) utilization (can individuals and households make good use of the food to which they have access?). These dimensions matter in a cumulative manner: if food is not available in an area, a household or individual will not have an adequate diet. But even when food is available, households must be able to afford to buy sufficient food. Even then, do they choose a nutritious diet or not? Do they prepare it properly to deliver its full nutritional value? Are they healthy enough to metabolize and absorb the nutrients? See Barrett (2010) for further discussion.

How do food security measures relate to monetary poverty measures? At extreme levels of poverty, monetary poverty and food insecurity are likely to be highly correlated, as affording a minimum number of calories is the first objective households and individuals will satisfy.

continued next page

Box 3.2 *(continued)*

This correlation is not perfect, in that a very poor household may receive social or humanitarian assistance and thus be in monetary poverty but not food insecure. As incomes increase, there are competing uses for the new income. It could go to more calories, more nutritious calories (for example, protein and micronutrients), or tastier calories, but it could also go to rent, utilities, health care, education, clothing, and other essentials. The important difference is that food security is just one (albeit important) dimension of household welfare, whereas monetary welfare (even without examining nonmonetary poverty) reflects the (in)ability to afford multiple dimensions of well-being. In addition, an increase in food prices might not result in a reduction in calories if households spend more to eat, but this results in a drop in nonfood consumption (and therefore welfare); in this case, food security measures will not pick up but monetary poverty will (in theory).

Food security and monetary poverty measures may differ for other reasons. For example, as Headey and Ecker (2012) observe for Ethiopia, the urban-rural gap in monetary poverty can be much smaller than for nutrition-based measures (such as child stunting), reflecting the difficulties of pricing subsistence consumption across geographies, unobserved seasonal shortfalls, and access to the quality essential services needed for nutrition, all of which are not well-captured in monetary poverty measures. Thus, monetary poverty is not necessarily the best proxy for food security. Headey and Ecker (2012) conclude that dietary diversity measures are the best food security indicators, both in times of stability and in times of shock, with monetary poverty measures second, ahead of other food security measures. Consequently, food security measures are best used when food security is the program or policy objective, while monetary poverty measures are best used when a broader welfare definition is being considered. This conclusion is reinforced by evidence from Jensen and Miller (2011), who find that in-kind food assistance or food subsidies have limited impact on food consumption (and thus food security) as households substitute their own spending away from food to nonfood essentials, behavior that is captured under monetary poverty measures.

This book focuses on the monetary dimension of welfare because it has greater implications for targeting. To determine the type of program and targeting objectives needed for a given dimension of multidimensional poverty, such as a child not being in school, means assessing why the child is not in school. Is this for financial reasons (cost of tuition and books, opportunity cost of working, and so forth)—in which case, a monetary poverty–targeted program might make sense—or for availability reasons

(there is no school nearby, the road to the school is impassable, or the teacher is often absent)—in which case a geographic or facility-targeted program might make sense. However, in this context, this book focuses more on methods to assess monetary poverty, since in many places non-monetary dimensions of poverty are directly observable while the monetary dimension is less so (and measurable food security indicators are well-established). This is not to discount the importance of also measuring and targeting other dimensions.

Even a monetary measure of welfare is not a straightforward concept in practice. Basic welfare economics starts with the notion that consumption is what provides utility (happiness). Consumption equals income plus any change in assets (net wealth), and these are concepts used widely in both the analytics of poverty and eligibility determination in social assistance practice. Consumption is expected to be somewhat less variable than income, buffered precisely by a change in assets—the proverbial savings for a rainy day, although of course assets can take forms much more varied than changes in cash savings and include the value of insurance, whether publicly provided or market based.

Measuring income, consumption, or assets is a somewhat inexact art, with all sorts of data issues (see, for example, Mancini and Vecchi [2020] for a recent update on the classic paper by Deaton and Zaidi [2002]). Indeed, the practicalities of what is measurable often trump the conceptual discussions of what it is desirable to measure, a topic taken up in chapter 6 in much more detail. This chapter describes some of the challenges in measuring these three concepts.

The body of survey work and poverty diagnostics contains much discussion about which measure of welfare to use, for which purpose, and when and operates with some "stylized facts." Each of these facts is intuitive and backed by survey evidence.[15] Similar issues are pertinent for direct eligibility determination processes since they are often quite like abbreviated household surveys, whether done in the applicant's dwelling or an office setting.

Income

- Income may be easily measured for households with a regular, stable payment (whether in cash, check, or e-transfer) from a factory or firm. This type of income is a number that people are likely to know and does not vary from payment to payment. It may be slightly more difficult to measure income for the increasing share of formal sector service workers working part time with variable hours per week, although the numbers will be salient enough that over short periods, a worker may recall them, and they may have to do the accounting to figure out an annual tax statement and thus be exposed in a salient way to that figure.

- It can be difficult to measure income when it comes largely from small-holder agriculture or informal home enterprises. It is likely that business accounts are intertwined with household spending, there is no written accounting, and revenue and expense streams are highly variable from day to day or month to month.
- Thus, income is easy to measure in the same places it may be easy to verify from paystubs or tax payments. In more informal settings, it is difficult for the household to put a number on income and for administrators to verify it.

Consumption

- Consumption may be easier to measure where it is focused on core food goods and a few other essentials, like water, fuel, rent, and purchases that are salient due to the difficulty in making ends meet. Consumption can be difficult to measure where it is spread across a wide variety of goods and high enough that each individual purchase may not be very salient to the purchaser.

Assets

- Assets may be hard to measure for several reasons as valuation of holdings may be difficult, especially when the asset portfolio of a household is diversified and/or when markets are thin.
- Financial assets, like money in stocks or savings accounts, may only apply to better-off households and can be even harder to ask about than income or consumption because people do not wish to reveal their wealth.
- Debt balances (credit card debt or debt to moneylenders) may also be hard to ask about, for reasons of stigma.
- It may be easier to get people to report that they own land or buildings or large durable goods like cars or machinery that are sources of pride and already probably known to their neighbors, but those assets may be hard to value.
- Livestock is an important asset for the poor in many countries, but it is changeable in number (through the births, slaughter, or other deaths of the animals) and subject to unit price fluctuations.
- For poor people, it may be difficult to measure some important assets because their values are so small that they may be classified as consumption items in surveys. Examples include stocking up on foodstuffs or purchases of small implements or inputs for the household enterprise, such as a new pick for a vegetable garden or a supply of fabric for a dressmaker.

Poverty Assessments and Eligibility Determination Mechanisms

In poverty assessments, income and consumption are commonly employed welfare measures used to classify households from the poorest to the richest. The differences in economic patterns and ease of measurement of household welfare just discussed have meant that income has tended to be used more in household surveys and distributional studies in richer countries. Consumption tends to be the core welfare measure in poorer countries, although the pattern is not absolute.

For eligibility determination, there is some flexibility in measuring welfare. Means testing focuses on income or some combination of income and assets. Inferential methods like hybrid means testing, proxy means testing, or small area poverty mapping associate observables with income or consumption.

Assets play a role in several kinds of eligibility determination mechanisms but in different ways. In means testing, holding assets of certain types or above a threshold value is often used to exclude households from eligibility ("affluence testing"). In hybrid means testing, the flow of consumption from productive assets may be imputed and that value added to more directly measured flows of income from wages and transfers. In proxy means testing, assets are used to help predict consumption or income, but final eligibility is determined based on the prediction only. In programs geared to respond to natural disasters, change in assets may trigger eligibility.

Where Should Eligibility Thresholds Be Set?

It is well acknowledged in the poverty literature that there is not an absolutely clear or unique place to draw the poverty line. Welfare distributions are continuous and often relatively flat over several deciles and/or there may be a sizable share of households just above the official poverty line who are not much better off than those just below it. The case of Indonesia illustrates this phenomenon (figure 3.2). While 10 percent of Indonesians were poor in 2018 according to the local poverty line, a further 18 percent of the population lived just above the poverty line but by less than 50 percent more.[16]

In part, poverty analysis deals with the problem of where to draw the poverty line by thoughtful grounding for an initial choice of lines, often by using a variety of poverty lines and careful interpretation. Usually, the authorities take due care in setting the initial anchoring of the basic poverty line, often in the cost of a food basket calculated to represent the expense of a low-cost, culturally acceptable, and nutritionally adequate food basket. This may be used as an "extreme" poverty line or "food" poverty line. The main or national poverty line that is often drawn tops up the basic poverty

Figure 3.2   Indonesia: Poverty and Near-Poverty
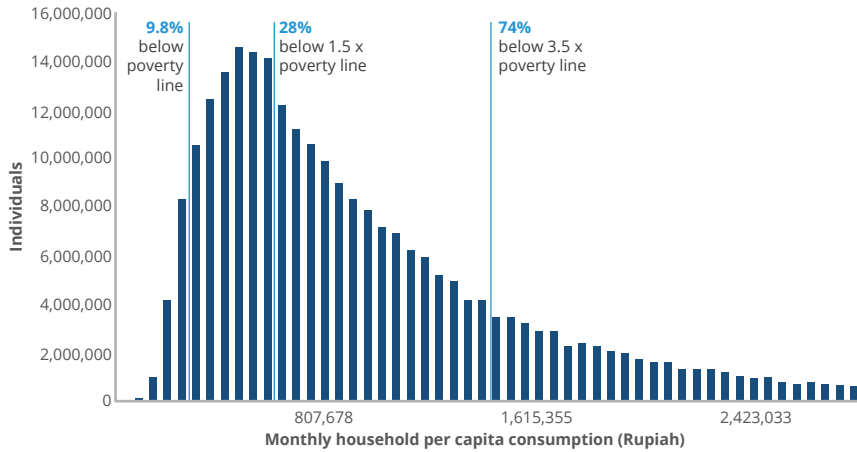
Source: Indonesia, National Social Economic Survey (SUSENAS) 2018.

line with some sort of calculation or allowance for nonfood expenditures. Often this is done by examining the nonfood expenditures of a "reference group" or households living around the estimated line. While the procedure is, roughly speaking, intuitive, it involves a host of details: Which foods in which combination? Which nutritional requirements to take into account? Only calories? Proteins? How many of a long list of micronutrients? Which vector of foods supplies those at low cost, which may differ in different agroclimatic zones? How much allowance should be made for flavor and culture over just nutrients? How to allow for nonfood expenditures—with a similar list of necessary goods and prices? If so, which goods? Or with a "share of expenditure" calculation? If so, which calculation? None of these questions has absolutely clear answers, much less answers that are precisely comparable across countries or time. Trying to answer them all results in widespread acceptance that there is a measure of arbitrariness in whatever poverty line is picked (see Haughton and Khandker 2009; Ravallion 1998; World Bank 2018a).[17] However, despite the fact that the construction of a national poverty line is part science and part art, it nonetheless represents a country's stated goal for the minimum standard of living to which it aspires for all its residents. As such, the national poverty line provides an anchor for determining the coverage of different social protection programs and against which eligibility can be assessed.

It is similarly common to use multiple thresholds for different programs in the social protection system. This reduces a bit the problem that there is no clear or unique place where people start/stop needing support.

Through reducing benefits from individual programs as welfare rises or layering multiple programs with different thresholds, it is possible to minimize stark cliffs in benefits. The use of multiple programs that are thoughtfully layered or segued together is common. This approach is desirable as families and individuals in need must have access to a packet of services to help them prosper enough that they are not economically vulnerable. In most countries that have moderate or high coverage for the overall social protection system, such layering of different services and programs is available to different segments of the population. For example, the nonpoor may receive assistance with natural disasters, catastrophic health insurance, or unemployment insurance. Those vulnerable to poverty may benefit from those programs plus seasonal assistance for utilities in cold climates or rainfall insurance for farmers. The chronic poor may receive all these plus basic income support to help provide for basic needs. Use of benefits differentiated by welfare level is somewhat less common. Of course, they are inherent in guaranteed minimum income programs and found sometimes in block rather than continuous form in the degree of subsidy for health insurance or utility price subsidies; they are less frequent in other forms of simple cash transfers.

The operational thresholds for individual programs may be set higher than the conceptual target group to ensure that errors of exclusion in the conceptual target group are reduced, although of course with an increased budget requirement. Errors of exclusion are always of concern, and simulations and targeting assessments usually show that those related to mis-measurement are particularly dense closer to the eligibility threshold.[18] Setting the threshold on the high side can help ameliorate the problem.

Figure 3.3 illustrates the results of simulations that were done to inform the government of an upper-middle-income country as it was considering compensation for a reduction in energy subsidies. The government's goal was to protect the poorest 30 percent of the population from the average income loss resulting from higher energy bills. The government was considering a hybrid means test to determine eligibility for the program. Simulations of such a scenario showed that with a threshold set at the third decile, a program could cover three-quarters of the group, but a quarter would be excluded. Then the analysts ran a simulation using a threshold covering 50 percent of the population. In this variant, coverage among the bottom 30 percent increased to 88 percent of the group. Of course, the cost increased as well in proportion to the larger coverage, by 67 percent. Therefore, an intermediate variant was considered that drew the threshold at the fifth quintile but tapered benefits to those in the fourth and fifth quintiles. This scenario maintains the higher coverage of the population of interest but reduces the extra costs to about 25 percent. Chapter 5 presents in detail the steps in setting up and interpreting such simulations, which

Figure 3.3   Simulations of Alternate Eligibility Thresholds for Coverage
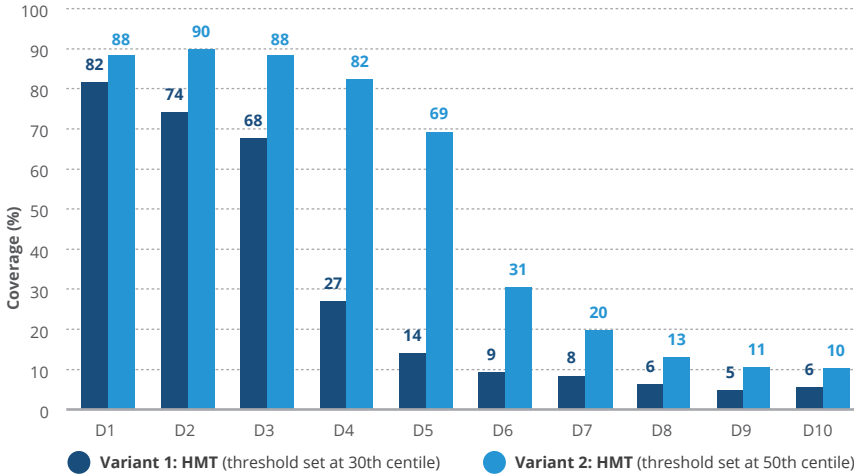
Source: Original figure developed for this publication.
Note: D = decile; HMT = hybrid means test.

are important for selecting eligibility assessment methods, eligibility thresholds, and benefit design.

Part II: Dynamics Add Some Further Challenges

How Dynamic Is Welfare?

The dynamics of welfare and poverty are considerable. In Africa, one-third of the population is persistently poor, while another third moves in and out of poverty (Dang and Dabalen 2019). In the Middle East and North Africa, chronic poverty accounts for around 50 percent of total poverty in the region (Dang and Ianchovichina 2018). In Jordan, in 2010, 14.4 percent of the population was poor in that their average annual per capita consumption was below the national poverty line. However, a further 18.6 percent of households fell below the line in at least one quarter of the year while having annual average consumption above the line (World Bank 2016). Figure 3.4 illustrates the welfare dynamics in selected East Asian countries over two-year panels, showing the significant degree of movement between different consumption groups (World Bank 2018b). The dynamics of welfare explain why vulnerability to poverty (as per the definitions in box 3.1) and the dynamics of social protection are such important topics.

Figure 3.4   Short-Run Mobility across Economic Classes, Selected Countries

a. Lao PDR, transition matrix, 2010–12

b. Indonesia, transition matrix, 2008–10

c. Philippines, transition matrix, 2007–09

d. Vietnam, transition matrix, 2010–12

Extreme poor   Moderate poor   Economically vulnerable   Economically secure   Middle class
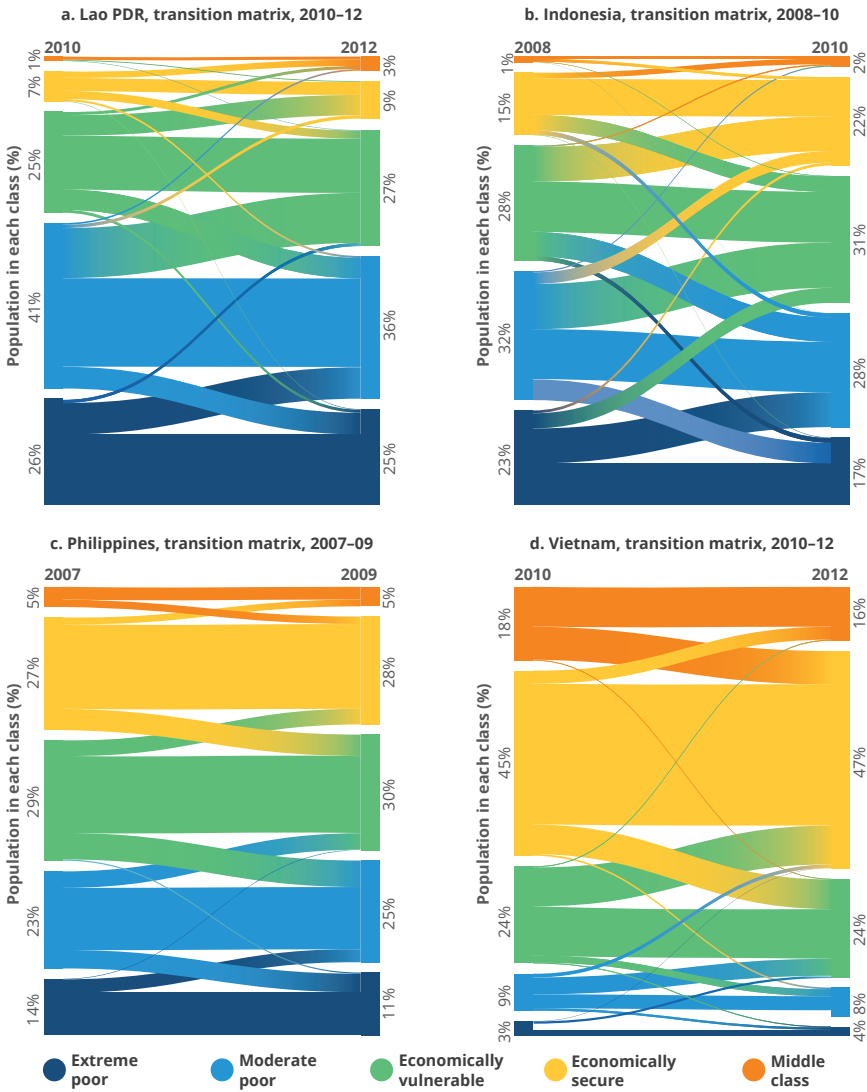
Source: Based on Indonesia, National Social Economic Survey (SUSENAS) 2018.

Because of these poverty dynamics, a higher income or consumption threshold can be estimated that defines a level of economic security. In a range of regional reports, the World Bank defines "middle-class" house-holds as those that are "economically secure"—that is, they have a low probability of falling into poverty in the next period based on their current

income level.[19] However, this income level represents the level at which households on average are unlikely to fall into poverty in the next period; many households with higher incomes will nonetheless be poor in the next period because they subsequently suffer a significant loss of income.

Variability in income is expected for many livelihoods. Farmers, agricultural workers, and the linked enterprises and communities may have good years and bad depending on the weather. Workers in the tourist industry work more hours per week or days per month "in season," as defined by the local climate or the school year in the country of residence of the tourists. Construction may be seasonal, limited during the rainy season or winter. Part-time shift workers may get more or fewer hours depending on variations in demand. In advanced countries, changes in labor markets have increased the number of workers with short-term, part-time, or gig work, with increased variability of income. In addition, individuals are affected by idiosyncratic job losses, illnesses that affect earnings, and changes in household structure that affect needs.

Where to draw the line between "normal variability" in income and "shocks" is not well defined. For some natural phenomena, there are relatively well-agreed thresholds. For example, droughts are commonly measured using the Standardized Precipitation Index, which categorizes rainfall amounts according to their standard deviation (SD) from historical norms.[20] Mild dryness (less than 1 SD) occurs every 3 years, moderate dryness (1.0-1.49 SD) every 10 years, severe dryness (1.5-1.99 SD) every 20 years, and extreme dryness (2 SD or greater) every 50 years. This measure can be used on both short timescales (soil moisture) and longer ones (groundwater and reservoir storage). Similar objective measures exist for flood risk (see map 3.1, later in this chapter)[21] and earthquakes.[22] Such well-defined scales do not exist for household-level shocks.

The ability of families to tolerate variability in income depends on both the magnitude of the variability and the household's initial position. To put it metaphorically, a man standing on tiptoes with just his nose above water can be swamped by a rise in water level of just a few inches. Another man only up to his knees in water can tolerate a much larger rise before being in danger. Hence, the policy actions to be taken depend on the starting point and the size of the losses.

The ability of families to tolerate shocks also depends on how widespread they are. Idiosyncratic losses can be minor or major for the individual household, but even when they are significant, the fact that they are scattered may make it easier for the household to cope. If an individual loses a job, there may be others available on short notice, or the family may be able to get by in the meantime with income from other members, savings, or a bit of help from extended family, friends, or neighbors. When losses are covariate, the severity may be compounded. When a whole industry suffers, the number of displaced workers may exceed the number of other

jobs that are easy to find, so reemployment will be harder, and extended family, friends, or neighbors may have lost their jobs too and be unable to help. If a tree falls on one family's roof, materials and help should be readily available to fix it. When a whole province suffers from a hurricane, many people need to fix their roofs at once and materials and skilled workers may become scarce and expensive.

Figure 3.5 illustrates the various scenarios that shocks can present. Individual A was poor before the shock but becomes poorer. A change in benefit levels—customized as in a guaranteed minimum income or with top-up benefits for ongoing social assistance recipients after a covariate shock—is appropriate. Individual B was not poor before the shock but is so after the shock. This is the classic case of the "vulnerable" or "new poor" and illustrates the need for programs that can expand to take in new clients as needed. Individual C represents someone who was quite well-off before the shock and suffers a very large loss but remains above the poverty line. In the ideal, an actuarially fair (commercial or self-financing) insurance program would be a good policy option. Individual D represents someone who was well-off before the shock and not much affected—such scenarios are a low priority for public policy action. The case of individual E is a reminder that not all shocks are negative, welfare may improve over time as well as deteriorate, and programs also need to consider when individuals should exit.

With its two functions of poverty reduction and risk management, social protection may aim to serve both the poor (chronic or new) and those who suffered large losses although they remain above the poverty line. In concept, social assistance is more suited to the poverty reduction function and the various types of insurance are more suited to the risk management function. However pure prefinanced insurance usually does not cover all who may be affected, and so there is often postdisaster public action financed with general tax revenues to compensate partially for losses, even among the nonpoor.

Figure 3.5   Stylized Effects of Shocks
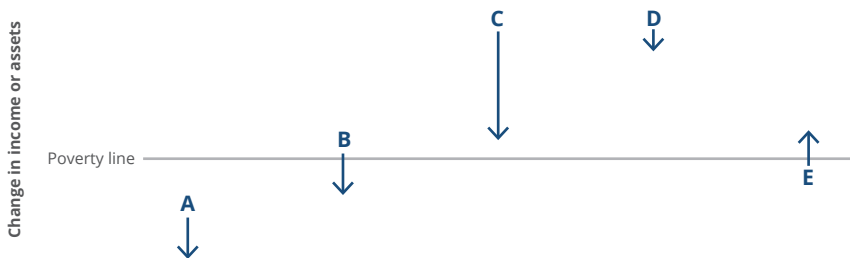
Source: Original figure developed for this publication.

Social Assistance Program Tools That Handle "Normal" Variability in Welfare

One part of handling variability in welfare is through giving weight in the overall social protection system to programs that offer insurance or quasi-insurance, or variable benefits over knowable cycles of differential need. Health, disability, and unemployment insurance are of course well known if not yet universally available parts of social protection. Various other programs can help compensate for the variability of earned incomes. For example, weather insurance can help reduce income variability for farmers in rainfed agriculture. Seasonal programs can help bridge seasonal fluctuations in needs. In cold climates, for example, home heating is necessary and may be a significant expense; thus, heating allowances are common. In poor countries, public works programs may operate in the agricultural slack season as a way to boost income in a lean period. Although it is less common, it would be simple to give larger payments for child allowance programs or conditional cash transfers at the beginning of the school year to help households handle the implicit (uniforms and shoes) or explicit (fees and school materials) costs of schooling. The whole know-how of these programs and their relative weight in social protection systems is beyond the scope of this book, but several of them have eligibility requirements that depend on welfare, and the operation of such programs reduces the variability in need that must be dealt with by other programs.

One part of handling the variability of welfare is via well-calibrated design choices and administrative processes to ensure that the programs actually include all that they are meant to include. The programs that are best able to handle income volatility are those that have open enrollment, budgets and budget flexibility so that they can fully fund all those who are found to be eligible, and eligibility determination mechanisms that are sensitive to changes in welfare. Since such programs represent a small minority of social assistance practice, it is important to consider what can be done in situations where the information, budget, or delivery system is unable to provide full flexibility. In these cases, one or more of the following approaches may be appropriate:

- Programs to address chronic poverty can use measures of longer-run welfare, like assets or consumption, in their eligibility criteria.
- Eligibility thresholds should be set considering vulnerability, income volatility, and the purpose of the program, with various plausible scenarios.
  - If the program can discern the characteristics of the most chronically poor,[23] for example, the bottom blue band in figure 3.4, then the program could be geared to those profiles.

○ The inverse approach would be to set thresholds relatively high so that the frequently poor households who happen to be having a good year or season when assessed are not excluded. Such a threshold could be set using the economic security approach. Returning to figure 3.4, a significant share of those in extreme poverty in period 1 rise to moderate poverty in period 2, but few rise above that, and some of those who were initially in moderate poverty fall into extreme poverty subsequently. So even if the threshold/group of deepest concern is the extreme poor, a program that covers the moderately poor might be needed. Of course, there will be volatility at the higher threshold, so a program that considers moderate poverty would not entirely solve the entry/exit problem (and makes the program more expensive), but it does put the threshold at a part of the income distribution where the harm of not receiving support during a bad period would be somewhat less.

○ Sometimes extra information may allow more precise ways of handling vulnerability. This approach tries to identify particular individuals or households above the poverty line that are vulnerable because their welfare is more susceptible to shocks than average because of where they live or the work they do. This approach differs from the average economic security approach by accounting for household and local characteristics to identify vulnerability in addition to estimated income. This is explored in more detail in chapter 6.

- An implication of income volatility is that entrance to programs should be continuously available, to allow anyone experiencing a dip in welfare to get timely support. This challenge of "dynamic inclusion" is frequently unmet for reasons of insufficient administrative practices (see chapter 4 and Lindert et al. [2020]) and contributes to errors of exclusion.

- Income volatility adds to the call for adequate, flexible, or entitlement-based budgets. When program enrollments are rationed, those who suffer a drop in welfare will not find room in the beneficiary roster even if they are assessed as eligible. Being on a waiting list is scant help.

With variability in welfare also comes the need to consider whether benefits should last for a defined period followed by automatic exit, or whether to require periodic recertification of eligibility. In several cases, time limits rather than recertification fit the logic of the program: a training program lasts as long as its curriculum takes to impart. Age-based child allowances grant benefits until the child ages out. Many emergency programs provide one-off or time-limited benefits that are presumed to match roughly with the degree and period of heightened need. Some programs have time limits that fit less well but are used nonetheless, as a way to ration (rotate) slots when the budget is too small to serve all who are otherwise eligible, or for fear of generating dependency on the benefit. In these cases, there is a high

risk that an automatic exit rule will generate errors of exclusion, and some people may remain or again become poor after the period of benefits is exhausted. For programs meant to provide ongoing benefits but only to those below a welfare threshold, recertification is a necessity.[24] This is discussed further as part of the delivery chain in chapter 4.

The appropriate periodicity of recertification may vary by context or group. Welfare may change faster in a thriving and transforming economy than in a stagnant one. Welfare may change faster for some groups than others. For example, people living in urban areas may have more dynamism in employment, compared with those in rural or remote locations who may remain dependent on agriculture with its cycles of income variability. Young workers trying to find their first steady jobs may experience greater volatility in income than older workers with more established jobs or livelihoods. The extreme poor are usually more likely to be chronically poor. Recertification periods could therefore be customized for different subgroups, which would also help reduce data collection costs.

Comparison of a handful of flagship programs shows the value of dynamic entry and exit. All the programs are major poverty-targeted cash transfers that rely on social registries for eligibility determination; three operate with survey sweeps of fairly long periodicity; and the others require on-demand registration and thus staggered and more frequent recertification. Not updating the list of beneficiaries for several years will result in a larger share of beneficiaries whose welfare has improved since intake but who remain in the program consuming scarce budgetary resources that could be allocated to the new poor. Although the administrative complexity of updating household circumstances and recertifying beneficiaries is the usual reason for infrequent recertification, the evidence on administrative costs in chapter 2 suggests that such costs are relatively small and many times smaller than the size of both inclusion and exclusion errors. Over time, the fact that the pool of beneficiaries does not accurately represent the intended pool will reduce the usefulness of the program for addressing chronic poverty or inequality and during large covariate shocks (through simple but effective measures such as benefit top-ups). Moreover, the lack of updates for a long period aggravates the political economy issue of recertification, by increasing the number of beneficiaries who have to be taken off program rolls at a given moment.

Evidence from countries with relatively long periods for recertification shows how targeting outcomes can deteriorate between rounds as the pool of beneficiaries gradually diverges from the originally intended group. For example, three countries—the Philippines, Pakistan, and Colombia—went through a period with long gaps in recertification. At the time of writing, all these programs have written new rules and initiated mass recertification programs due to such concerns. The Philippines plans to carry out more frequent recertification[25]; Pakistan is moving to a mixed model[26]; and

Colombia is preparing to move to a fully on-demand system with program-specific recertification requirements.[27]

• *The Philippines.* The current list of the Philippines' conditional cash transfer program, Pantawid Pamilyang Pilipino Program (4Ps) is drawn from the first wave of registration for the social registry (Listahanan 1), which was conducted in 2009.[28] The share of families in the poorest quintile, as measured by periodic household surveys, declined from 75 percent in 2009 to 53 percent in 2013, to 46 percent in 2017 (Acosta and Velarde 2015; Velarde 2018; World Bank 2020e).

• *Pakistan.* Eligibility for Pakistan's flagship Ehsaas Kafaalat (previously known as the Benazir Income Support Programme [BISP] unconditional cash transfer) program was determined via the National Socioeconomic Registry, which was established in 2011 based on a national census sweep using the proxy means test–based poverty scorecard. Since then, the welfare of some beneficiaries has improved, but there was no mechanism to enroll those who fell newly under the thresholds (figure 3.6). An administrator-driven registry update started in 2016, and in 2019, on-demand initiatives were incorporated.[29] As of May 2021, the BISP completed 90 percent of the expected national caseload, and enrollment based on the new data was underway.

Figure 3.6   Welfare Transitions among Benazir Income Support Programme—BISP Beneficiaries
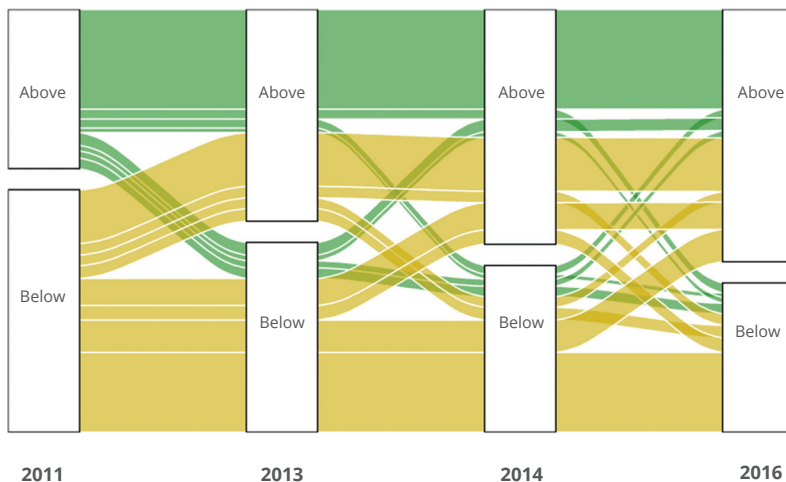
Sources: DIME 2020; surveys by Oxford Policy Management.

- *Colombia.* On a more compressed timetable, Colombia shows similar issues, although it has some provisions for on-demand entry and updates.[30] In 2015, four years after the big wave of recertification of the System for the Selection of Beneficiaries for Social Programs, as many as 65 percent of the households in the database were mistakenly identified as poor. The well-being of those households had improved enough between 2011 and 2015 to reclassify them as nonpoor, but the system was still categorizing them as poor. Meanwhile, 17 percent of the cases saw a deterioration in their situation between 2011 and 2015, but they were still classified as nonpoor (DNP 2016[31]).

Countries that rely on open-ended eligibility with frequent recertification and automated updates supported by data interoperability show more stable targeting outcomes:

- *Brazil.* Brazil has a fully dynamic registration process and two-year recertification cycle for its Bolsa Familia conditional cash transfer program. It has shown much more stable targeting outcomes: during 2015–18, coverage of the poorest quintile remained in the band of 57–60 percent each year, and the share of benefits accruing to the poorest quintile has risen somewhat, from 55 to 63 percent.[32]
- *Chile.* The social registry in Chile is fully dynamic (see box 5.7, in chapter 5). The eligibility threshold for the noncontributory child allowance (Subsidio Único Familiar, SUF) is set at 60 percent of the population.[33] The registry's statistics show fairly stable targeting outcomes: coverage of the poorest quintile was 36 or 37 percent in 2011, 2013, and 2015 and declined to 30 percent in 2017. The share of benefits accruing to the poorest quintile has risen somewhat, from 44 to 50 percent over the same period.[34]
- *Armenia.* The family benefit and social benefit programs in Armenia offer on-demand eligibility, with recertification every 12 months.[35] At recertification, the applicant must resubmit an application declaration and the necessary documents, except the information on income, which is updated on a quarterly basis (the household/family member submits a reference from employer[s]). Beneficiary incidence remained relatively stable over 2010–18, with 50 to 60 percent of the benefits accruing to people in the poorest quintile, with the most progressive results in the most recent years. The program has low coverage overall, about 13 percent of the population and 35 percent of the poorest quintile.
- *Turkey.* The social assistance programs in Turkey combine open eligibility with frequent information updates, which trigger re-estimation of the proxy means test. The programs use the Integrated Social Assistance Service Information System, a social registry that is interlinked with 24 administrative databases. These databases automatically supply about

40 percent of the information on household circumstances used to score households, and the system updates the administrative records automatically at least every 45 days. The rest of the household circumstances are updated at least once a year, through a household visit. The updates result in a continuation of status, a change in benefit, or exit from the program. The beneficiary incidence of unconditional cash transfers remained relatively stable over 2010–08, with about 70 percent of the benefits accruing to people in the poorest quintile, although that share dropped to 65–62 percent in the most recent years. The program has low coverage overall, at about 6 percent of the population and 21 percent of the poorest quintile.[36]

Too frequent recertification that requires significant private costs from applicants can result in exclusion error, which in turn hinders program effectiveness. For example, in 2018, Albania's Ndhima Ekonomike social assistance program for the poorest operated a three-month recertification cycle. An operational audit of the program was carried out when the program switched its eligibility from hybrid means testing to proxy means testing. The audit found that about 5 percent of the beneficiaries were removed from the program after three months, at recertification, because they failed to produce all their justificative documents in time. Moreover, poor understanding of program rules, such as the need to declare their employment and car ownership, which were subsequently cross-checked automatically with the respective registries, resulted in the rejection of additional applications before they were scored, which was equivalent to about 2.4 percent of the beneficiaries. Such high administrative barriers artificially contributed to a reduction in Ndhima Ekonomike's caseloads during 2018. Subsequently, the program has improved its outreach efforts (Honorati 2019).

It may be acceptable for the dynamism of program rosters to be a bit lower than the variability of income. Even with fully open access, a family facing a dip in income may not bother to apply for social assistance if the setback puts them just a bit below the eligibility line, especially if they expect that the dip will be short-lived or the benefits for which they would qualify would be low because they are barely eligible. Conversely, various decisions about income disregards, benefit tapers, recertification periods, or the minimum duration of benefits can allow households with incomes that modestly exceed the eligibility thresholds to remain on benefits for some period.

In addition, when eligibility periods are longish and income dynamics are high, the accuracy of targeting social assistance programs in assessments based on household surveys must be cautiously interpreted, as welfare at the time of eligibility determination may be rather different from welfare at the moment of the household survey for a large share of households.

This does not necessarily indicate inaccurate eligibility determination at the moment of entry into the program, although it does imply that resources would be less focused on the poorest when the survey is done.

Large Covariate Shocks: Modification of Programming and Systems to Determine Eligibility and Benefits

The toll of climate change and natural disaster, pandemic, conflict, and economic shocks on poverty and prosperity is clear. Climate change could result in an additional 100 million people living in extreme poverty by 2030 (Hallegatte et al. 2016). A fifth of the world's population, 1.5 billion people, live in areas of high flood risk (World Bank 2020f; see map 3.1). COVID-19 is expected to increase global poverty for the first time since 1998, with current estimates on the order of 90 million to 120 million people (World Bank 2020f).[37] Depending on the length of the pandemic, estimates suggest that half the rise in poverty could be permanent, and by 2030, the poverty numbers could still be higher than the baseline by 60 million people.[38] Moreover, conflict remains an enduring problem for many parts of the world. The 37 countries formally classified as fragile, in conflict, or suffering from violence make up only 10 percent of the world's

Map 3.1   Share of the Global Population with High Flood Exposure

IBRD 45219 | AUGUST 2020
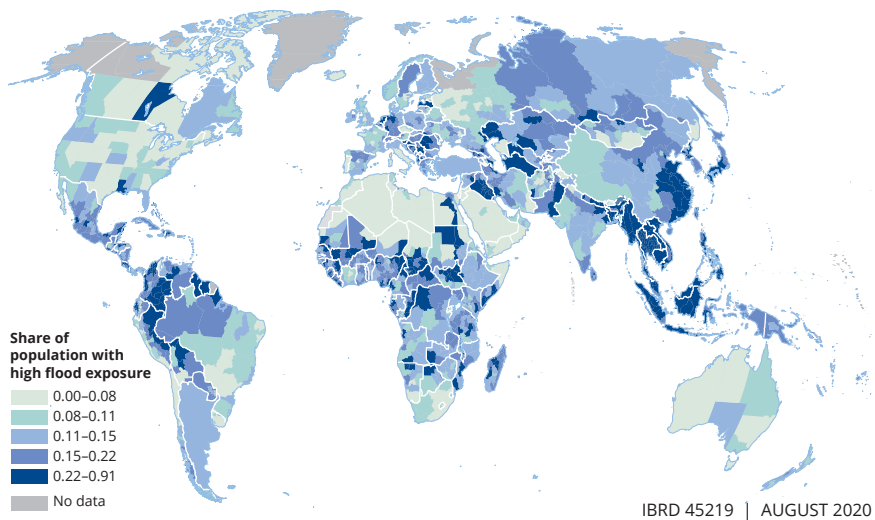
Source: Rentschler and Salhab 2020.

Note: The share is the percentage of the population in a given territory or principal administrative division, in which 20 percent of all territories and divisions are shown.

population but more than 40 percent of the global poor. Conflict and poverty are often co-located in Africa (World Bank 2020f). Hence, without proper planning, large social protection systems, and flexible programs and systems, shocks can offset most of the past years' gains against poverty, as measured at the global extreme poverty rate.[39]

When trying to forestall the long-term consequences of shocks, the speed of assistance can be of utmost importance. The logic is intuitive. If assistance is to prevent a negative coping tactic, it must be timely, before a family's baby becomes malnourished, before the family withdraws its children from school, marries off a child bride, sells its assets, racks up high-interest debt, or loses its home, workshop, or land. Each of these coping tactics can be very difficult to reverse, ratcheting down the individual's or family's welfare for years or the rest of their lives (Báez, Fuchs, and Rodríguez-Castelán 2017; Caruso 2017; Hallegatte et al. 2017; Hill, Skoufias, and Maher 2019).

Linking program action or expansion to ex-ante determined triggers can help speed policy responses. Expanding program mandates and budgets for responding to shocks triggered by indicators determined ahead of time can facilitate the speed of response. Such indicators can include those that are pertinent for natural disasters, such as windspeed, flood stage, rainfall, or vegetative conditions. Indicators of economic covariate shocks can also be identified, such as the Consumer Price Index, food price increases of a certain size or duration, and unemployment increases of a certain size, above a certain level, or for an extended duration. Chapter 4 provides some examples. These triggers may determine that coverage or benefit levels should change, or that the decision about which specific households would benefit should rely on other mechanisms for eligibility determination.

Each natural disaster and economic crisis renews attention to the "wicked problem" of providing emergency social protection, especially to the informal sector. The problem has several dimensions. Some dimensions have fairly well-identified technical solutions, but they have not yet been fully implemented, especially in poorer countries. Other dimensions of the problem are tougher to solve.

• One part of the problem is the *incomplete coverage of social assistance even for the chronically poor in normal times*. In most crises, the already poor also suffer and it is critical to help them. Top-up benefits are by far the simplest and fastest social protection response to implement, especially where payment mechanisms are electronic. These responses use precrisis targeting on the understanding that people in need in "normal" times will have limited margins to cope with shocks. Moreover, good coverage of basic social assistance in normal times usually implies development of the basic building blocks of delivery systems, which are also important to further crisis response. So, increasing coverage in normal

times provides an important tool for crisis response. Some sorts of social assistance in noncrisis times are especially suited for building resilience and thus reducing problems in the face of emergencies.[40]

- One part of the problem is that an increase in the client base also implies *a need for incremental funding*. Whether arranged through entitlement-based authorizations, contingency funds with preset triggers, disaster-related reinsurance schemes, or executive action to authorize short-term budget reallocations or overruns, financing needs to be quickly available. When the crisis is big, adequate response often implies that funding needs are so large as to require legislative action or recourse to humanitarian appeals, but as these are usually slow, it is desirable to have some access to initial funding quickly. This is not directly a problem of how to determine eligibility, but it shapes issues around how narrowly or broadly a response can be targeted and whether programs will be dynamic, adjusting to need, or static, delimited by the budget.

- One part of the problem is the increase in the client base means that *people who were not previously needy will become needy and must be able to access new programs in a hurry*. This puts a premium on "open access" or "on-demand" application processes or emergency field work, which are topics taken up in chapter 4 on delivery systems.

- Another part of the problem is the *incomplete coverage of foundational ID systems* that expedite linkages among existing information bases and facilitate rapid electronic payments. IDs are already on the development agenda for many reasons, with many countries working quickly to improve the coverage or quality of their systems. The importance of IDs for social protection and crisis response adds weight to the agenda and is also taken up in chapter 4.

- One part of the problem is *incomplete financial inclusion* impedes convenient payment mechanisms. Digital financial services are facilitated by the tech revolution and the availability of digital foundational IDs for know-your-customer rules, but significant regulatory and market-building issues may need to advance before low-cost accounts are common for the bottom of the pyramid. Well-done digital social protection payments can speed response, lower transaction costs for clients, and improve governance, but care must be taken that the digital divide does not lead to errors of exclusion.

- *Sudden changes in incomes for the many in the informal sector* are an inherently difficult problem, even in countries that have largely solved the issues listed above. This is especially true for those who are not usually poor, whose income is not observed through formal channels like payroll or income tax registries or easily proxied through poverty maps or characteristics of their dwelling. There is no magic solution. Over time, more developed interoperable information systems and access to new

sources of data may help, but it is among the most intransigent problems, and countries must implement active policies to seize the benefits of the secular improvements in technology and data (see chapter 6 for more).

• And of course, just when more is being asked of them more quickly, emergency responses often have to deal with *disruptions in delivery systems*. In natural disasters, this may be due to interruptions in power, internet, and transport; possibly displacement of client populations; and destruction of some social assistance offices. During the COVID-19 crisis, pandemic control put a premium on solutions that did not require face-to-face interaction or queues at social assistance service points.

The recurrence of shocks and crises and the premium on swift response raise the challenge of how social protection systems can prepare in advance. The various dimensions of improving shock responsiveness in social protection systems include resilience-building programs and improving the shock responsiveness of data and information systems, finance, and institutional arrangements and partnerships (Beazley, Solórzano, and Barca 2019; Bowen et al. 2020). Given the focus of this book on eligibility determination, we focus on this element of the whole. Several of these topics pertain to improvements in or adaptations to delivery systems, which are treated in chapter 4. Chapter 5 discusses the pros and cons of different targeting methods with respect to emergency response, and chapter 6 reviews how to implement them. The remainder of this chapter focuses on some of the wider policy considerations.

There is always a balancing act for shock response programs on whether to focus on the already poor, the new poor, or the most affected. The issue has angles of both values/moral philosophy and pragmatics. In some cases, the conceptual tension may not be empirically large because welfare and susceptibility to shocks are often inversely correlated. The poorer are more likely to live on marginal lands that are more subject to flooding, landslides, droughts, or high temperatures. The poorer are likely to have less resilient housing or means to evacuate even when early warning systems give notice. The poorer are more likely to work in agriculture affected by weather. The urban poorer are more likely to work in the informal sector and thus are not covered by labor-related protections or insurance when the economy falters. Thus, those who are most affected by shocks and crises may be the already poor or nearly poor. For example, Bottan, Vera-Cossio, and Hoffmann (2020) show how COVID-19–related losses in employment were highest at the bottom of the income distribution. If disaster-related effects are not measured in absolute terms but as a share of baseline assets or income, the share of losses by the poor may be as much as double that of the nonpoor (Hallegatte et al. 2017).

Assisting the poor during or after a crisis by expanding existing social assistance is a high priority and likely one of the easiest policies even when the situation calls for a wider response. Helping the chronic poor after a crisis may not be sufficient, but it is likely to be pertinent to the crisis and/or a way to repair a hole in social policy. The chronic poor may most quickly have to resort to negative coping tactics. The simplest and fastest social protection response is to issue an emergency top-up payment to people who are already in some social assistance program. Often expanding ongoing but low-coverage programs is the next fastest option. When existing registries include waiting lists or information on households above the normal cut-off, it can be done almost by the stroke of a pen. In other cases, it may take handling new applications, but the existence of procedures and staffing may help speed that work.

A crisis response beyond helping the already poor may be called for to reach the new poor or those who have suffered significant losses. As COVID-19 was declared a pandemic, many countries issued various levels of travel restrictions, stay-at-home orders, or quarantine orders, which quickly reduced economic activity to various degrees depending on the strictness and enforcement of the orders and the degree to which work and commerce could be conducted remotely. According to the high-frequency telephone tracking surveys[41] supported by the World Bank and its client country partners, across countries, 62 percent of households reported income loss. An average of 34 percent of the respondents reported stopping work, 20 percent of wage workers reported lack of payment for work done, 9 percent reported job changes due to the pandemic, and 60 percent of households receiving remittances before the pandemic reported a decline in remittances. The widespread and deep losses underscored the need for responses that reached above the usual poverty line or eligibility thresholds for programs for the chronic poor.

Often the benefit structure for social assistance crisis response programs is simple, aiming for wide coverage among those affected, with flat or only minorly customized benefits. Such a structure has three advantages:

- The simplicity allows greater speed and transparency in delivery. Quantifying losses precisely requires detailed household-by-household assessment, which would likely be slow and possibly imprecise.
- The breadth helps reduce errors of exclusion. There may be less tolerance for errors of exclusion (and more for errors of inclusion) in times of disaster and crisis as disasters tend to increase the feeling of solidarity and reduce the "othering" of the poor and vulnerable to poverty that may be present in times of prosperity. The crisis may also diminish, at least temporarily, concerns over work disincentives if work is hard to come by or less productive than usual.

• Setting flat benefits to a level that covers a reasonable share of the costs of the poor or vulnerable, and thereby a lower share of the costs for better-off households, responds to a social welfare function that values assistance to the poorer more than that to the less poor. It may also mitigate concerns over using public funds to assist the nonpoor who have suffered losses, as it is likely to be assisting them with only a portion of their losses.

But the flat benefit structure has the disadvantage of not matching benefits closely to losses. The "going big" coverage, even where losses have not been measured, will imply substantial budget and may not be sustainable over time. Thus, a successively greater degree of scrutiny, narrower eligibility, or smaller benefits may be brought to bear after the initial moment of crisis.

Four country examples illustrate how this can work:

• Pakistan's response to the 2010 flood shows how targeting can be initially simple and broad and refined in the second phase. The 2010 flood covered a fifth of the country and affected 20 million people. Pakistan's Citizen's Damage Compensation Program implemented a flat benefit structure in two phases, reaching 1.6 million households. Households in phase I of the program were selected using geographic targeting—residents of villages where more than 50 percent of the housing stock was damaged qualified. This implied obvious errors of inclusion and exclusion, but it allowed for a quicker response, especially as flood waters remained for many weeks and carrying out a house-by-house assessment was very difficult in three of the four affected provinces. For phase II, benefits were again flat, but the eligibility criteria were modified. A simple household-by-household housing assessment was conducted, with third-party verification of a sample and exclusionary filters, such as international travel, to exclude the higher income groups. Extra considerations ensured that households headed by women and people with disabilities would be registered. This increased scrutiny allowed more precise targeting but took longer (World Bank 2013). Kosec and Mo (2017) show that the provision of these transfers not only restored livelihoods and replaced damaged assets, but also had an enduring effect by easing mental burdens and thus raising aspirations for the future among the beneficiary population, which helped to address the negative effects of natural disasters on people's expectations.

• Brazil demonstrates a case of scaling down by reducing benefits. Faced with the COVID-19-related crisis in 2020, the government mounted a multipronged approach, including establishment of the Emergency Assistance Cash Transfer Program to protect informal sector workers and the unemployed in April 2020. The initial threshold for coverage and eligibility criteria were high and simple; by the end of its first month of

implementation, 60 million Brazilians had benefited, eventually reaching 68 million. The initial benefit was set at R$600 (US$114) for up to two eligible adults in families without a formal income. After five months, the benefit dropped to R$300 (US$57) for a further five months. From April 2021, the benefit amount dropped to R$250 (US$47.5)[42] and the "assistance unit" was changed to the household, dropping the expected number of beneficiaries paid from 68 million to 47 million.[43] The emergency program ended in October 2021, but assistance has continued to a large portion of the Emergency Assistance beneficiaries through an expanded version (17.6 million families, up from 14.6 million) of the long-standing Bolsa Familia program now called Auxílio Brasil, with its traditional means testing[44] (World Bank 2020a; Folha de São Paolo;[45] de Arruda et al. 2021).

- Bhutan represents a case of both refining targeting criteria and reducing benefits as a crisis unfolds. In response to the COVID-19 pandemic, the Royal Government of Bhutan launched the Druk Gyalpo Relief Kidu.[46] The objective was to provide support to those who were directly affected, including employees and self-employed workers in tourism and tourism-dependent sectors, other affected sectors, and returnee migrants. The program offered a direct cash transfer to individuals or loan interest support for employers.[47] The program was implemented in phases of three months to provide flexibility to applicants and the government as the impacts of the pandemic on economic opportunities varied among sectors and over time. From April to June 2020 (phase I), the direct cash transfer component set benefits at Nu 12,000 (US$160 in April 2020) (mainly for tourism and tourism-dependent sectors and returnee migrants) or Nu 8,000 (for other sectors), plus Nu 800 per child per month to eligible beneficiaries with children. Since then, the main benefits dropped to Nu 10,000 and Nu 7,000, respectively, and the eligible sectors are regularly revised as economic recovery unfolds. Eligible applicants receive the benefits from the month in which they apply, within each three-month cycle. On average, over 75 percent of the applicants in each phase reapplied in the subsequent phases,[48] and a large increase in new applications occurred in August and December 2020 when the country went into full lockdown. However, the number of reapplicants has been decreasing (for example, 67 percent of the phase I applicants reapplied in phase II, while only 53 percent of the phase II applicants reapplied in phase III), reflecting economic recovery and the fact that some of the selected sectors (such as taxi drivers and restaurants) for a given phase were excluded from the eligibility criteria in the subsequent phase.
- Canada represents a case where an emergency program carried the load of the response until changes in permanent programs could be put in place. Canada took a fast and easy approach with the Canadian

Emergency Response Benefit for workers who lost their jobs due to COVID-19 shutdowns. The benefit offered Can$500 per week, eliminating the proportionality of the traditional unemployment insurance and requirements for job search but enabling payout within a very short period. Most payments flowed within a week of application[49] and rolled out finally to 9 million Canadians, about a quarter of the population.[50] The flat and relatively high benefit allowed certain groups with no significant shortfall, such as young people living with their parents or part-time workers, to qualify and even to receive more than their lost employment income. Several tweaks were made in subsequent weeks after the launch, including wage subsidy programs to help counteract work disincentives and allow part-time workers to benefit. But the main parameters remained in place until the end of September 2020 when the Simplified Unemployment Insurance program (modified to allow greater coverage than previously) and three related recovery schemes for the self-employed, caregivers, and those needing extra sick leave were designed to taper the support and reinstitute more traditional return-to-work incentives (Godbout 2020).

Confronted with the scale of (expected) impacts of quarantines and stay-at-home orders on work and incomes, governments quickly took decisions to expand social protection. By the end of March 2020, three weeks after the World Health Organization declared COVID-19 a global pandemic, 84 countries had taken social protection measures, and within three months, 195 countries had declared social protection policy responses. A year after the crisis began, the World Bank policy tracker identified more than 3,000 planned or implemented responses in 212 countries covering about 2.3 billion people. The responses covered the gamut of social protection instruments, but cash transfers and waivers or subsidies to utilities were among the most common. With the range of countries and programs used in the response, the targeting practice was diverse but there were some noteworthy facets:

- Sixty-eight countries topped up existing cash transfer programs with additional benefits, thus riding on prior targeting decisions and their implementation for at least part of their response.
- One hundred sixty-six countries initiated new, mostly temporary programs, averaging four months in duration, which meant that targeting practice had to be unusually swift but any errors of inclusion would be a short-lived call on resources.
- In many countries, COVID-19 responses have been more broad-based than much of the existing social assistance or responses to prior crises. A few headline-grabbing responses were even universal (for example, Mongolia, Timor-Leste, and Tuvalu), and several more reached more

than two-thirds of the population (for example, El Salvador and Morocco) or very large absolute numbers of people (Brazil, China, India, Indonesia, Pakistan, and the Philippines). For a subset of developing countries, it is possible to compare coverage pre–COVID-19 using data from the Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE) with the emergency response. In these cases, the COVID-19 response approximately doubled coverage in nonfragile countries and increased 10-fold in some fragile states due to their low initial coverage (figure 3.7), a pattern that was first measured in the December 2020 tracker[51] and revalidated with minor changes in the May 2021 issue.

- To accomplish targeting so quickly, programs often drew on existing data, sometimes in novel ways or via rapidly developed electronic outreach and application processes (see more discussion in chapters 4 and 6).

Assessments of impacts are beginning to emerge. In the United States, early responses were so big and fast that initially poverty actually fell by 2.3 percentage points, although employment rates fell by 14 percent in April 2020—the largest one-month decline on record (Han, Meyer, and Sullivan 2020). In Brazil, policy responses were large enough (with simple estimations that do not model any behavioral change) that the poverty rate declined from 12 percent in 2019 to 8 percent while the temporary program was in place, and the difference between poverty rates for blacks and whites also dropped significantly and despite a drop in labor force

Figure 3.7   Increase of Coverage due to COVID-19 Responses
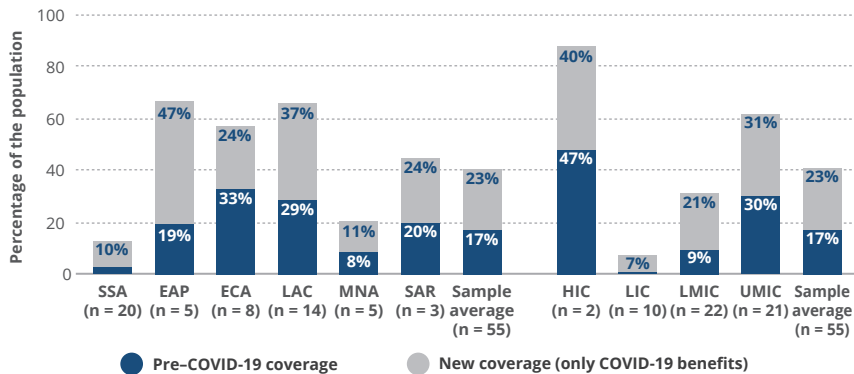
Source: Gentilini et al. 2020.

Note: EAP = East Asia and Pacific; ECA = Europe and Central Asia; HIC = high-income countries; LAC = Latin America and the Caribbean; LIC = low-income countries; LMIC = lower-middle-income countries; MNA = Middle East and North Africa; SAR = South Asia; SSA = Sub-Saharan Africa; UMIC = upper-middle-income countries.

participation of 7 percentage points (Menezes Filho, Komatsu, and Rosa 2021). The more common finding for low- and middle-income countries was that the scale of losses was larger than could be compensated by the social assistance response, resulting in widespread food insecurity and/or reductions in employment and income (Beazley, Bischler, and Doyle 2021; Egger et al. 2021). Of course, the social protection response helped to ameliorate the damage to a degree where it was substantial enough. (See, for example, Gallego et al. [2021] for an evaluation of Colombia's Ingreso Solidario, Abay et al. [2020] for the impacts of Ethiopia's rural Productive Safety Net Program for those who benefited, and Cho et al. [2021] for the Philippines Social Amelioration Project, phase 1.)

Crises often foment progress in social protection systems, through both quick responses and increased action in the sector in the years following the crisis. The following years can be quite important, due to the often increased appreciation of the importance of social protection, the lessons concerning the deficiencies in systems the crisis revealed, and the know-how generated in response. The 1997 East Asia financial crisis moved forward consensus on the importance of social assistance not only in the region but beyond. The APEC (2001) Lessons of Crisis document was subscribed to by the Asian and Latin American member countries as well as the International Monetary Fund, World Bank, Inter-American Development Bank, and Asian Development Bank. It articulates lessons that remain familiar and have been restated with more or less elaboration after the other major international rounds of crises—that the keys to crisis response are (1) preparedness, (2) a range of instruments to be used according to the context, and (3) a range of concerns to be addressed. A decade later, the food and fuel price crisis marked a turning point in the development of poverty-targeted cash transfers, especially in Africa. Although they were small, many programs were launched or expanded in the years of high food prices, and several of the programs were the predecessors or starts of eventually much larger programs (Beegle, Coudouel, and Monsalve 2018; Grosh et al. 2011). The sector's development since the last round of crises set the basis for response to the COVID-19–induced economic crisis. Many of the programs had quick top-up benefits, and many more of the ministries used the tools they had built into social registries and payment mechanisms to implement COVID-19 response programming even where it was in stand-alone, one-off programs.

This is playing out again, with countries considering or deciding to expand or improve their precrisis social assistance, inspired by their COVID-19 crisis responses. Some countries are changing procedures and rules to improve coverage and flexibility. For example, India has undertaken several changes: it has revised the rules for its State Disaster Response Funds to allow up to half the funds to be used for social protection responses

to "notified emergencies." It has legislated a new Social Security Code and the rollout of a Social Security Board, which consolidates formerly fragmented systems and focuses administrative effort to improve coverage of informal sector workers by key insurance and assistance programs. To improve coverage of migrants, India has taken several actions, premier among them making access to the flagship Public Distribution System portable (World Bank 2020b, 2020c). Some countries are using the heightened expectations of social protection to expand the coverage and programming of their permanent programs. Morocco is a case in point (box 3.3), although other countries, such as Brazil, Nigeria, and Pakistan, have also announced plans or begun actions to increase coverage of their programs by some degree. A number of countries will continue their move to digital registration and payment (see more in chapter 4), since once the channels are built, they remain available for further use.

Box 3.3

Morocco's Progress toward Universal Social Protection

In response to the COVID-19 crisis, Morocco is expanding its social protection system, from a low-coverage, low-adequacy, and regressive precrisis system. The COVID-19 crisis hit the country in its incipient stage of reforming costly and weak social assistance. Despite Morocco's significant spending on social assistance, equivalent to 3 percent of gross domestic product (compared with an average of 1.8 percent among other middle-income countries), half of this spending financed regressive consumer subsidies for liquefied petroleum gas, sugar, and flour.[a] At the same time, the noncontributory social assistance system was fragmented, with more than 40 programs of modest generosity and signficant coverage gaps. The two most important programs before the crisis, the conditional cash transfer program Tayssir[b] and the noncontributory health insurance fee waiver, the Medical Assistance Plan (RAMED),[c] still left about half of the poorest quintile uncovered and offered relatively low benefits.

Morocco's response to the unprecedented COVID-19 shock was fast, resolute, and comprehensive. The government of Morocco rapidly mobilized human, financial, and technological resources to reach a large share of the population affected by the strict confinement policy, including its poor and vulnerable population. For informal workers, who lack social security and are not affiliated with the main pension

Box 3.3 *(continued)*

house (CNSS), the government set a specific budget line under its Special Pandemic Fund CAS-COVID-19 called Tadamon (meaning "solidarity" in Arabic) to distribute an emergency cash transfer.[d] The program provided a subsistence amount to applicants based on their family size: (1) DH 800 (about US$82) for families of one or two members, (2) DH 1,000 (about US$103) for families of three or four members, and (3) DH 1,200 (about US$124) for families of five or more members. Tadamon's coverage was large, reaching two-thirds of all households in the country.[e]

Postcrisis reforms. On July 29, 2020, the king of Morocco announced a broad set of measures to support the recovery of the Moroccan economy and the resilience of households, including a deep-reaching reform of the social protection and labor sector,[f] which were futher detailed in a blueprint by the Ministry of Finance[g] and legislated as the Framework Social Protection Law in September 2021. To increase the resilience of households, especially those in the informal sector, in the first phase, Morocco will expand and roll out mandatory health insurance until the end of 2022 and a new universal family allowance until December 2024. The second phase, to be implemented from 2024 to 2025, will expand the old-age pension coverage and compensation for loss of employment to those who do not already have it.

Source: World Bank (2020d).
a. For example, only 7 percent of the liquefied petroleum gas subsidy benefited households in the bottom quintile of consumption, while about 50 percent went to the richest quartile.
b. Tayssir supports poor and vulnerable families, conditional on their school-age children attending school, with the aim of improving school enrollment and reducing dropout rates. It was first launched in 2008 and is managed by the Ministry of National Education, Vocational Training and Scientific Research. During the 2019/20 school year, Tayssir disbursed about US$200 million to support 2.4 million children in primary and secondary school. The program pays the household DH 60 (about US$7) per month for every registered child in the first and second grades; DH 80 (about US$8) per month for children in the third and fourth grades; DH 100 (US$11 equivalent) for children in the fifth and sixth grades; and DH 140 (about US$15) per month for children in the seventh, eighth, and ninth grades (*secondaire collégial*).
c. As of September 2019, RAMED covered about 14.4 million individuals.
d. The website is www.tadamoncovid.ma.
e. The first wave of the emergency cash transfers (ECTs) in April 2020 reached 4.2 million households. Eligible households automatically received a second wave in May 2020. Through the grievance and redress mechanism, 1.25 million households were added at a later date. Therefore, the first and second ECT waves benefited a total of 5.45 million households, of which 2.8 million were RAMED and 2.65 million were non-RAMED recipients. Further, female-headed households comprised around 1.17 million, constituting 21 percent of the first and second waves of the ECTs. A third wave was distributed in July 2020 and covered all previous beneficiaries. As of October 2020, the ECTs had distributed a total of DH 16.150 billion (US$1.65 billion).
f. The king's speech can be accessed at https://www.moroccoworldnews.com /2020/07/313044/full-text-of-king-mohammed-vis-21st-throne-day-speech.
g. The minister's presentation can be accessed at https://www.finances.gov.ma /Publication/cabinet/2020/Pr%C3%A9sentation%20Minstre_03.08.2020_VFr.pdf.

What it takes to prepare social protection systems to be shock responsive overlaps heavily with the move to universal social protection for stable times, with multiple programs that address different population needs and characteristics. Improved coverage, dynamism, program design and flexibility, and social protection delivery systems are vital for both crisis and noncrisis times. Shock responsiveness puts a premium on flexible programs, agile delivery systems, and real-time information and encourages the development of the social insurance subsector as well as social assistance. In particular, greater access to more information and big data underpins much of this. A greater capacity to manage various large administrative databases can lead to better delivery and targeting of programs. How this is done in times of shocks depends on the nature of the shock but helps drive dynamic assessment.

In summary, this chapter was a journey from the general but abstract idea of "helping the poor and vulnerable" through many of the complexities involved in reaching implementable definitions and program designs. The first half of the chapter covered defining welfare, the assistance unit, and eligibility thresholds. The second half showed the considerable importance of dynamics, as idiosyncratic shocks affect families and natural disasters or economic crises affect nations. Handling these calls for a huge range of policy decisions, programs, and operational capacities. In normal times, this means the following:

- Defining objectives and the assistance unit for each program
- Defining the welfare measure of choice for each program
- Understanding local volatility of incomes
- Setting eligibility thresholds considering vulnerability and the purpose of each program
- Determining the periodicity and eligibility threshold or criterion of each program's recertification.

To make systems responsive to shocks, this also means the following:

- Preparing systems ahead of covariate shocks to facilitate the speed of response
- Designing emergency responses that can cover the informal sector
- Balancing focus on the already poor, the new poor, and those most affected
- Balancing broad coverage and flat benefits against a more tailored design.

Chapters 4, 5, and 6 elucidate how different aspects of the move from abstraction to implementable concepts can be done. Chapter 4 looks at how these considerations are reflected in the design and operational readiness of the delivery chain. It considers each step of the chain: outreach, intake and

registration, assessment of needs and conditions, eligibility and enrollment decisions, determination of benefits and services, notification and onboarding, and provision of benefits and services. In particular, it highlights that "targeting failures" can happen at any step of the chain and discusses the roles of IDs, social registries, payment mechanisms, and adaptive social protection.

Chapter 5 relates the concerns raised in this chapter to the choice of targeting method. It first observes the evolution of targeting practice over the past two decades. It then outlines a framework for choosing method(s), accounting for fit-for-purpose with respect to program objectives and feasibility in different country contexts. Each method is assessed in the light of this framework, and quantitative simulations are introduced to help select methods.

Chapter 6 looks at notions of data and inferences and how they connect to some of the options among the methods. The chapter reviews the nature of data for inferring eligibility. The rest of the chapter provides an in-depth, technical overview of each of the methods considered—geographic, means testing, hybrid means testing, proxy means testing, and community-based testing—considering best practices, how big data and new inference methods influence them, and how they can be adapted for shocks.

Notes

1. For example, the United Nations Household Survey Capability Program defines a household as a group of people who live together, pool their money, and eat at least one meal a day together (United Nations 1989). Eurostat defines a household in the context of surveys on social conditions or income, such as the European Union Statistics on Income and Living Conditions or Household Budget Survey, as a housekeeping unit or, operationally, as a social unit: having common arrangements, sharing household expenses or daily needs, or in a shared common residence (https://ec.europa.eu/eurostat/statistics-explained /index.php/Glossary:Household_-_social_statistics#:~:text=A%20 household%2C%20in%20the%20context,in%20a%20shared%20 common%20residence).
2. An adult equivalence scale is an adjustment made in calculating household welfare that accounts for its demographic composition, with the underlying hypothesis that people of different ages have different needs. For example, if food is the largest item in consumption and children need fewer calories than adults, they would have a weight less than one, in proportion to established scales of caloric need by age and sex. Economies of scale attempt to capture the idea that two people together can live more cheaply than two people separately. For example, it takes less than twice the fuel to heat a larger cooking pot, housing will be less than twice as large, and consumer durables such as a television or refrigerator can be shared.

3. The use of a different unit for analysis than for policy can confuse evaluation. In evaluations of targeting outcomes, facile use of household survey data could show errors of inclusion or exclusion when no eligibility decision process was wrongly conducted. Decisions about the appropriate assistance unit bear deep policy consideration and evaluative data analysis that matches. In some countries, care is taken so that the definitions of assistance units and households can be mapped to one another, with the social assistance program working within the nomenclature of the national statistical office.

4. Marriage rates have been declining in many countries, especially in the West, with the share of never-married women ages 40–44 rising, for example, in Australia (from 5 percent in 1986 to 15.6 percent in 2006), Brazil (from 9 percent in 1980 to 33.8 percent in 2010), France (from 7.5 percent in 1985 to 27.9 percent in 2009), and the United States (from 4.8 percent in 1982 to 13.8 percent in 2012). In 2010, more than 40 percent of births in France, Norway, and Sweden were to women in cohabiting relationships, compared with 25 percent in the United States (Besharov and Gilbert 2015). The share of women ages 20–24 in informal unions in Latin America rose between the 1970 and 2000 censuses to rates between 23 (Mexico) and 69 (Peru) percent (Esteve, Lesthaeghe, and Lopez-Gay 2012). In Jamaica, according to 2014 data from the Registrar General's Department, 85 percent of children are born outside formal marriage (Gleaner 2016).

5. See, for example, Lambert, van de Walle, and Villar (2017); Premand, Schnitzer, and van de Walle (forthcoming); Trócaire (2017); and van de Walle (2013).

6. In Mali and Niger, eligibility for the cash transfer is based on the welfare of the compound household of husband, all wives, children, and others. The behavioral change elements that accompany the cash transfers that focus on protecting and boosting children's human capital outcomes while empowering women are individually targeted to the mothers. In both countries, men are invited to participate in the behavioral change elements as well. In Ethiopia, in contrast, the rural Productive Safety Net Program treats each wife and her respective children as a separate household, with the husband ascribed as a member of the first wife's household.

7. Here we focus on informal fostering where one or both parents remain prominent and trusted deciders of a child's living arrangement as opposed to legal fostering where the state has terminated parental rights due to abuses or neglect of duty.

8. Demographic and Health survey data from 16 African countries show that the percentage of households with a foster child ranges from 15 percent in Ghana to 37 percent in Namibia (Vandermeersch 1997). Lloyd and Desai (1992) use the same survey data to calculate the percentage of children living away from their biological parents and find rates ranging from 5 percent in Burundi to 28 percent in Botswana.

9. The share of 18- to 29-year-olds living with their parents since the COVID-19 crisis began has reached 52 percent, surpassing the previous peak of 48 percent during the Great Depression. (https://www.pewresearch.org/fact-tank /2020/09/04/a-majority-of-young-adults-in-the-u-s-live-with-their-parents -for-the-first-time-since-the-great-depression/).

10. The threshold for block pricing can be difficult to set in ways that distinguish well the poorer from the less poor. For example, in Jordan, 85 percent of the poorest decile of households have total electricity consumption within the subsidized levels, but so do 53 percent of the richest decile (Rodriguez and Wai-Poi 2020).

11. HelpAge, "A Resource for Promoting Dialogue on Creating a New UN Convention on the Rights of Older Persons," https://social.un.org/ageing -working-group/documents/Coalition%20to%20Strengthen%20the%20 Rights%20of%20Older%20People.pdf.

12. Kakwani and Subbarao (2005), writing when the HIV/AIDS epidemic was caus- ing a surge of concern over the unprecedented strain on families, investigated poverty rates by various family structures. They found that while some of the expected generalizations about elderly poverty were true for regional aggrega- tions, the magnitudes were markedly different. The poverty headcount in "skip- generation" households comprised of only elderly and children was about twice as high as all poverty in Côte d'Ivoire, but only half as high in Cameroon. The "elderly living alone" had a headcount poverty rate 8 percentage points higher than average in Uganda but 29 points lower in Nigeria. Households with both elderly and prime-age adults had slightly lower poverty headcounts in Burundi, Madagascar, and Uganda than all households, but in Côte d'Ivoire, The Gambia, Kenya, and Zambia, headcount poverty rates for households with elderly mem- bers were 10 or more percentage points higher than for all households.

13. Silwal et al. (2020) update the estimates with new surveys and similar meth- ods, but the summary note contains less detail. The headline numbers for 2017 were that 17.5 percent of children in the world (or 356 million) younger than 18 years lived on less than $1.90 in purchasing power parity per day, compared with 7.9 percent of adults ages 18 and older. The poverty rates of children at the $3.20 and $5.50 lines were 41.5 and 66.7 percent, respectively.

14. See http://hdr.undp.org/en/2020-MPI on the Global MPI and World Bank (2018a) on the World Bank measure as well as an extended discussion on multidimensional measures in general.

15. See The Living Standards Measurement Study (LSMS) website (http://surveys .worldbank.org/lsms/), which presents the household survey program by the World Bank's Development Data Group. Since its inception in the early 1980s, the program offers technical assistance to national statistical offices in the design and implementation of multitopic household surveys and the measure- ment and monitoring of poverty. Other seminal publications include Beegle et al. (2012); Deaton and Grosh (2000); Smith, Dupriez, and Troubat (2014); World Bank (2003).

16. The use of social welfare functions with continuous weights means it matters a lot less where the line is drawn in terms of measuring a program's impact on welfare. Chapter 2 shows that use of the poverty gap means that poor households who receive benefits that are inadequate to bring them over the line are still counted as being positively affected. More generally, the distribu- tional characteristic discussed in chapter 7 can be used to assess program per- formance where any household receiving a benefit is considered positively affected but with much higher weights on poorer households.

17. To facilitate comparability across countries, the World Bank has constructed a suite of international poverty lines that take into account the different costs of living in different countries ($1.90 per person per day for extreme poverty, $3.20 for lower-middle-income countries, and $5.50 for upper-middle-income countries), as well as multidimensional definitions of poverty that go beyond just the monetary dimension. Care in updating, interpretation, and looking at multifaceted pictures of poverty help to overcome the fundamental problem of definitions in poverty analytics. Moreover, as countries become wealthier and almost no households are poor when evaluated against any reasonable absolute poverty line, countries tend to move toward using a relative poverty line. For example, the European Union uses 60 percent of median income as its poverty line, which means some households will always be considered poor, regardless of how far above the absolute poverty lines their incomes are.

18. Errors of exclusion related to delivery systems (lack of outreach, difficulty in mustering required documentation, high transactions costs, and so forth) may be worst among the very poorest.

19. See López-Calva and Ortiz-Juarez (2011) for the original methodology. Household panel survey data are used to observe income/consumption in one period and then a subsequent period. Lowess or logit regressions are then used to determine the probability of a household being in poverty in the second period given their first period income/consumption. From this probability curve, an economic security line can be set, for example, the point at which a household has less than 10 percent chance of being poor next year. See Ferreira et al. (2013) for an application to Latin America and the Caribbean, Ruggeri Laderchi et al. (2017) for an application to East Asia and the Pacific, and World Bank (2019) for an application to Indonesia.

20. See, for example, John Keyantash and National Center for Atmospheric Research Staff (eds), "The Climate Data Guide: Standardized Precipitation Index (SPI)," https://climatedataguide.ucar.edu/climate-data/standardized -precipitation-index-spi, last modified August 7, 2018.

21. The measure defines 0 meter = no risk; 0–0.15 meter = low risk; 0.15–0.5 meter = moderate risk; 0.5–1.5 meters = high risk; over 1.5 meters = very high risk (World Bank 2020f).

22. Earthquake magnitudes are commonly measured by the well-known Richter scale. Under this scale, the magnitudes, effects, and frequency are (1) 2.5 or less, usually not felt, but can be recorded by seismograph, 900,000 per year; (2) 2.5 to 5.4, often felt, but only causes minor damage, 30,000 per year; (3) 5.5 to 6.0, slight damage to buildings and other structures, 500 per year; (4) 6.1 to 6.9, may cause a lot of damage in very populated areas, 100 per year; (5) 7.0 to 7.9, major earthquake with serious damage, 20 per year; (6) 8.0 or greater, great earthquake that can totally destroy communities near the epicenter, one every 5 to 10 years.

23. This should be done with country-specific empirical analysis. Hypotheses to test would be whether the chronic poor are among the poorest (because they have to increase their income the most to rise above the eligibility threshold) or those with barriers to increased earnings, such as low skills, poor health or

disability, residence in poor or remote areas, or membership in a historically excluded group by tribe, ethnicity, caste, religion, and so forth.

24. To counter concerns about disincentives around exit, there are various options. For example, to encourage beneficiaries to take new jobs or formal sector jobs, workers who do so may be able to retain benefits for a period of time, there may be some sort of income disregard that allows workers to keep part of incremental income, or there could be a minimum benefit period or a right to return without wait-listing should the workers' income again fall. To soften the transition and reduce political backlash, families found upon recertification to be above the exit threshold may receive benefits for a fixed transitional period, possibly at a lower level.

25. In the Philippines, legislation (the "4Ps Act" signed in April 2019) now mandates a regular revalidation of beneficiary targeting every three years. Enumeration for a new Listahanan 3 started in October 2019, was disrupted and delayed by COVID-19, and as of December 2021, data encoding and validation were at the final stage. Once completed, the database should be used for targeting and recertification of beneficiaries.

26. In 2020, Pakistan adopted a recertification strategy, through approval by the Benazir Income Support Programme (BISP) Board, for its flagship Ehsaas Kafaalat program. BISP relied on analysis of various sources of data, including nationally representative and specialized surveys and cross-checks between the old and new National Socio-Economic Registry (NSER) data, to inform the development of the strategy. The strategy covers both NSER and program-specific elements, including (1) introducing socioeconomic status bands for the NSER rather than a single cutoff to allow a broader range of programs to utilize the NSER as a social registry, (2) provisions for two quarters of transition benefits to all exiting cash transfer beneficiaries, and (3) continuity of benefits to exiting families with children enrolled in the education-linked conditional cash transfer program until primary school completion. The government is working on a plan that will allow multiple points of entry (physical registration desks at the subdistrict level, door-to-door surveys, and virtual services) (World Bank 2021a).

27. DNP CONPES 3877; World Bank 2021b.

28. Although the social registry was updated in 2015 (Listahanan 2), a substantial share of 4Ps households was missed, so the Department of Social Welfare and Development opted not to utilize the registry to recertify or exit families in 4Ps.

29. The government is working on a plan that will provide multiple points of entry (physical registration desks at the subdistrict level, door-to-door surveys, and virtual services) to allow data updates and program entry and exit on a more frequent basis.

30. Between periodic survey sweeps, to keep information up to date, households may request a survey for the first time or request that their household information be updated. These requests are made through the municipal SISBEN offices, which are run and funded by the municipalities themselves (World Bank 2021b).

31. https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3877.pdf.

32. Calculations based on ASPIRE data.
33. https://www.chileatiende.gob.cl/fichas/33112-subsidio-familiar-suf.
34. Calculations based on ASPIRE data.
35. The two benefits use the same targeting and eligibility system but differ on the target groups (families with or without children) and benefit levels (MoLSA, World Bank, and UNICEF 2021).
36. The 6 percent coverage refers to cash transfer beneficiaries, while 40 percent coverage refers to the social registry information system's information on poor/vulnerable households in general (nonbeneficiaries and beneficiaries of any social assistance in general). Some of the functions of Turkey's social registry, such as case management, graduation to work, and support for informal working poor, are still nascent. Robustly developing more frequent, comprehensive case management systems and options for informal vulnerable households in Turkey that integrate targeting and benefits between noncontributory social assistance and contributory social security systems, such as by subsidizing a unified health insurance system for all, will be especially needed for the green transition and improving the adaptability of social protection to shocks such as pandemics or earthquakes.
37. Homi Kharas, blog (2020), https://www.brookings.edu/blog/future-development/2020/10/21/the-impact-of-covid-19-on-global-extreme-poverty.
38. https://www.worldbank.org/en/topic/poverty/brief/projected-poverty-impacts-of-COVID-19.
39. The global extreme poverty rate is expected to increase for the first time since 1998, bringing it back to the level in 2017. The World Bank predicts that poverty will reach about 9.1–9.4 percent of population instead of declining to about 7.9 percent, as pre-COVID-19 estimates indicated (https://www.worldbank.org/en/topic/poverty/brief/projected-poverty-impacts-of-COVID-19).
40. There are many ways to build resilience to hazards ex ante, and resilience building has been a growing line of programming in social protection in recent years. Resilience building tends to be a "no regrets" policy as it often includes the same sorts of activities that would be important in reducing chronic poverty and building shared prosperity. At the community level, and within social protection, public works programs may help with landscape management and water retention in drought-prone areas, upgrade flood defenses in flood-prone areas, or provide landslide protection on densely settled, steep slopes. These policies can then benefit poor people who receive wages for temporary employment on the schemes and, along with their larger communities, may benefit from the protections provided in the works done. At the household level and within social protection, resilience building usually focuses on poor and vulnerable households that have slim margins between their baseline state and destitution. Activities that raise their incomes or diversify them in ways that lower risk and mechanisms that allow them to build assets or savings or join insurance pools all raise resilience. While it is very important for adaptive social protection, the topic of resilience building per se is broader than the focus of this book, which is targeting and eligibility determination.

41. "COVID-19 High-Frequency Monitoring Dashboard," World Bank Group, Washington, DC, http://www.worldbank.org/en/data/interactive/2020/11/11 /covid-19-high-frequency-monitoring-dashboard.
42. The single-person household benefit was set at R$150 (US$28.5), and the female-headed household benefit was set at R$375 (US$ 71.25).
43. https://www.cnnbrasil.com.br/business/2021/03/30/novo-auxilio -emergencial-lista-de-quem-tem-direito-deve-sair-nesta-quinta-1/.
44. Decree 10852, https://www.in.gov.br/en/web/dou/-/decreto-n-10.852-de-8 -de-novembro-de-2021-357706502.
45. https://www1.folha.uol.com.br/mercado/2021/08/bolsonaro-entrega-ao -congresso-mp-do-novo-bolsa-familia-e-pec-dos-precatorios.shtml.
46. www.royalkidu.bt.
47. Eligible loan borrowers also received a total interest waiver for three months, followed by a 50 percent interest waiver.
48. Applicants in phase III, which initially covered October to December 2020, did not need to reapply for phase IV to provide immediate relief during the second national lockdown.
49. The government also simplified enrollment processes, with eligibility checks done mostly through data matching or ex post (https://globalnews.ca/news /6804623/coronavirus-all-cerb-applications-approved/).
50. Tracey Lindeman, 2020, https://fortune.com/2020/10/23/canada-unemployment -cerb-economy-growth-coronavirus/.
51. The calculation is done for the countries with pre–COVID-19 coverage measured in ASPIRE. The baseline data should be taken as an upper bound estimate since administrative data tend to overestimate rates due to possible program overlaps.

References

Abay, Kibrom A., Guush Berhane, John Hoddinott, and Kibrom Tafere. 2020. "COVID-19 and Food Security in Ethiopia: Do Social Protection Programs Protect?" Policy Research Working Paper 9475, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34794.

Acosta, Pablo, and Rashiel Velarde. 2015. "An Update of the Philippine Conditional Cash Transfer's Implementation Performance." Policy Research Working Paper 104164, World Bank, Washington, DC. https://documents1.worldbank.org /curated/en/322971468178773885/pdf/104164-BRI-P082144-P150519-SP -Policy-note-No-8-PUBLIC.pdf.

Ainsworth, Martha, and Deon Filmer. 2002. "Poverty, AIDS, and Children's Schooling: A Targeting Dilemma." Policy Research Working Paper 885, World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle /10986/19289/multi0page.pdf?sequence=1&isAllowed=y.

Akresh, Richard. 2009. "Flexibility of Household Structure Child Fostering Decisions in Burkina Faso." *Journal of Human Resources* 44 (4): 976–97. https://www.jstor .org/stable/20648927?seq=1#metadata_info_tab_contents.

APEC (Asia-Pacific Economic Cooperation). 2001. "Social Safety Net in Response to Crisis: Lessons and Guidelines from Asia and Latin America." APEC, Singapore. https://www.apec.org/Publications/2001/02/Social-Safety-Net-in-Response -to-Crisis-Lessons-and-Guidelines-from-Asia-and-Latin-America-February.

Ardington, Cary, and Murray Leibrandt. 2010. "Orphanhood and Schooling in South Africa: Trends in the Vulnerability of Orphans between 1993 and 2005." *Economic Development and Cultural Change* 58 (3): 507–36. https://www.journals .uchicago.edu/doi/pdf/10.1086/650414.

Báez, Javier E., Alan Fuchs, and Carlos Rodríguez-Castelán. 2017. *Shaking Up Economic Progress: Aggregate Shocks in Latin America and the Caribbean*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/28892.

Barrett, Christopher B. 2010. "Measuring Food Insecurity." *Science* 327 (5967): 825–88. https://www.science.org/doi/10.1126/science.1182768.

Beazley, Rodolfo, Jana Bischler, and Alexandra Doyle. 2021. "Towards Shock-Responsive Social Protection: Lessons from the COVID-19 Response in Six Countries." Oxford Policy Management, Oxford, UK. https://www.opml.co .uk/files/Publications/A2241-maintains/maintains-towards-shock-responsive -social-protection-synthesis-report.pdf?noredirect=1.

Beazley, Rodolfo, Ana Solórzano, and Valentina Barca. 2019. "Study on Shock-Responsive Social Protection in Latin America and the Caribbean: Summary of Key Findings and Policy Recommendations." Oxford Policy Management, Oxford, UK. https://www.opml.co.uk/files/Publications/a1537-shock-responsive -social-protection-latin-america-caribbean/summary-of-key-findings-and -policy-recommendations.pdf?noredirect=1.

Beegle, Kathleen, Aline Coudouel, and Emma Monsalve. 2018. *Realizing the Full Potential of Social Safety Nets in Africa*. Africa Development Forum. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/29789.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson. 2012. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics* 98 (1): 3–18. https://econpapers.repec.org/article/eeedeveco/v_3a98_3ay_3a2012_3ai _3a1_3ap_3a3-18.htm.

Besharov, Douglas j., and Neil Gilbert. 2015. "Marriage Penalties in the Modern Social Welfare State." Policy Study No. 40, R Street Institute, Washington, DC. http://www.welfareacademy.org/pubs/family/Marriage_Penalties_in_the _Modern_Social-Welfare_State.pdf.

Bottan, Nicola L., Diego A. Vera-Cossio, and Bridget Hoffmann. 2020. "The Unequal Impact of the Coronavirus Pandemic: Evidence from Seventeen Developing Countries." Discussion Paper 785, Department of Research and Chief Economist, Inter-American Development Bank, Washington, DC. https:// publications.iadb.org/publications/english/document/The-Unequal-Impact -of-The-Coronavirus-Pandemic-Evidence-from-Seventeen-Developing -Countries.pdf.

Bowen, Thomas, Carlo del Ninno, Colin Andrews, Sarah Coll-Black, Ugo Gentilini, Kelly Johnson, Yasuhiro Kawasoe, Adea Kryeziu, Barry Maher, and Asha Williams. 2020. *Adaptive Social Protection: Building Resilience to Shocks*. International

Development in Focus. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/33785.

Brown, Caitlin, Rossella Calvi, and Jacob Penglase. 2018. "Sharing the Pie: Undernutrition, Intra-Household Allocation, and Poverty." https://ssrn.com/abstract=3199129 or http://dx.doi.org/10.2139/ssrn.3199129.

Brown, Caitlin, Martin Ravallion, and Dominique van de Walle. 2019. "Most of Africa's Nutritionally Deprived Women and Children Are Not Found in Poor Households." *Review of Economics and Statistics* 101 (4): 631–44. https://www.mitpressjournals.org/doi/full/10.1162/rest_a_00800.

Caruso, Germán Daniel. 2017. "The Legacy of Natural Disasters: The Intergenerational Impact of 100 Years of Disasters in Latin America." *Journal of Development Economics* 127: 209–33. https://www.sciencedirect.com/science/article/pii/S0304387817300317.

Cho, Yoonyoung, Jorge Avalos, Yasuhiro Kawasoe, Doug Johnson, and Ruth Rodriguez. 2021. "Mitigating the Impact of COVID-19 on the Welfare of Low Income Households in the Philippines: The Role of Social Protection." COVID-19 Low Income HOPE Survey Note No. 1, World Bank and Australia Department of Foreign Aid and Trade, Washington, DC. https://reliefweb.int/sites/reliefweb.int/files/resources/Mitigating-the-Impact-of-COVID-19-on-the-Welfare-of-Low-Income-Households-in-the-Philippines-The-Role-of-Social-Protection.pdf.

Dang, Hai-Anh H., and Andrew L. Dabalen. 2019. "Is Poverty in Africa Mostly Chronic or Transient? Evidence from Synthetic Panel Data." *Journal of Development Studies* 55 (7): 1527–47.

Dang, Hai-Anh H., and Elena Ianchovichina. 2018. "Welfare Dynamics with Synthetic Panels: The Case of the Arab World in Transition." *Review of Income and Wealth* 64: S114–S144. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3097066.

de Arruda, Pedro Lara, Marina, Lazarotto de Andrade, Silva Falcão Tiago, Diana Teixeira Barbosa, and Matteo Morgandi. "Auxílio Emergencial: Lessons from the Brazilian Experience Responding to COVID-19" (English). World Bank, Washington, DC.

Deaton, Angus. 2016. "Measuring and Understanding Behavior, Welfare, and Poverty." *American Economic Review* 106: 1221–43. https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.106.6.1221.

Deaton, Angus, and Margaret Grosh. 2000. "Consumption." In *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*, edited by Margaret Grosh and Paul Glewwe, 91–134. Washington, DC: World Bank. http://hdl.handle.net/10986/25338.

Deaton, Angus, and Salman Zaidi. 2002. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis." LSMS Working Paper 135, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/14101.

DIME (Development Impact Evaluation). 2020. "Improving BISP Household Targeting through Estimation of Poverty from Satellites, Facebook Advertising Data and Open Street Maps." World Bank, Washington, DC.

DNP (Departamento Nacional de Planeación). 2016. CONPES 3877. "Declaratoria de importancia estratégica del Sistema de Identificación de Potenciales Beneficiarios (Sisbén IV)." Bogota.

Egger, Dennis, Edward Miguel, Shana S. Warren, Ashish Shenoy, Elliot Collins, Dean Karlan, Doug Parkerson, A. Mushfiq Mobarka, Gunther Fink, Christopher Udry, Michael Walker, Johannes Haushofer, Magdalena Larreboure, Susan Athey, Paula Lopez-Pena, Salim Benhachmi, Macartan Humphreys, Layna Lowe, Niccolo F. Meriggi, Andrew Wabwrire, C. Austin Davis, Utz Johann Pape, Tilman Graff, Maarten Voors, Carolyn Nekesa, and Corey Vernot. 2021. "Falling Living Standards during the COVID-19 Crisis: Quantitative Evidence from Nine Developing Countries." *Science Advances* 7 (6): eabe09. doi:10.1126/sciadv.abe0997; https://www.science.org/doi/10.1126/sciadv.abe0997.

Esteve, Albert, Ron Lesthaeghe, and Antonio López-Gay. 2012. "The Latin American Cohabitation Boom, 1970–2007." *Population and Development Review* 38 (1): 55–81. https://www.jstor.org/stable/41857357?seq=1#metadata_info_tab_contents.

Ferreira, Francisco H. G., Julian Messina, Jamele Rigolini, Luis Felipe López-Calva, Maria Ana Lugo, and Renos Vakis. 2013. *Economic Mobility and the Rise of the Latin American Middle Class*. World Bank Latin American and Caribbean Studies. Washington, DC: World Bank. http://hdl.handle.net/10986/11858.

Gallego, Jorge, Bridget Hoffmann, Pablo Ibarrarán, María Paula Medina, Camilo Pecha, Olga Romero, Marco Stampini, David Vargas, and Diego A. Vera-Cossio. 2021. "Impactos del programa Ingreso Solidario frente a la crisis del COVID-19 en Colombia." Technical Note 2162, Departamento de Investigación y Economista Jefe, Inter-American Development Bank, Washington, DC. https://publications.iadb.org/publications/spanish/document/Impactos-del-programa-Ingreso-Solidario-frente-a-la-crisis-del-COVID-19-en-Colombia.pdf.

Gentilini, Ugo, Mohamed Almenfi, Ian Orton, and Pamela Dale. 2020. "Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures." World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/33635.

Gleaner. 2016. "Marriages Seeing A Decline: A Serious Threat to Public Health and National Development." *Gleaner*, May 5. http://jamaica-gleaner.com/article/news/20160507/marriages-seeing-decline-serious-threat-public-health-and-national-development.

Godbout, Luc. 2020. "Canadian Emergency Response Benefit (CERB): From Emergency to Recovery." *Perspectives on Tax Law & Policy* 1 (3). https://www.Ctf.Ca/Ctfweb/En/Newsletters/Perspectives/2020/3/200308.Aspx.

Grosh, Margaret, Colin Andrews, Rodrigo Quintana, and Claudia P. Rodriguez Alas. 2011. "Assessing Safety Net Readiness in Response to Food Price Volatility." Social Protection Discussion Paper 1118, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/26825.

Guven, Melis U., and Phillippe G. Leite. 2016. "Benefits and Costs of Social Pensions in Sub-Saharan Africa." Social Protection Discussion Paper 1607, World Bank, Washington, DC. http://hdl.handle.net/10986/24945.

Hallegatte, Stephane, Mook Bangalore, Laura Bonzanigo, Marianne Fay, Tamaro Kane, Ulf Narloch, Julie Rozenberg, David Treguer, and Adrien Vogt-Schilb. 2016. *Shock Waves: Managing the Impacts of Climate Change on Poverty*. Climate Change and Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/22787.

Hallegatte, Stephane, Adrien Vogt-Schilb, Mook Bangalore, and Julie Rozenberg. 2017. *Unbreakable: Building the Resilience of the Poor in the Face of Natural Disasters*. Climate Change and Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/25335.

Han, Jeehoon, Bruce D. Meyer, and James X. Sullivan. 2020. "Income and Poverty in the COVID-19 Pandemic." BPEA Conference Drafts, Brookings Papers on Economic Activity, Brookings Institution, Washington, DC. https://www .brookings.edu/wp-content/uploads/2020/06/Han-et-al-conference-draft.pdf.

Haughton, Jonathan, and Shahidur R. Khandker. 2009. *Handbook on Poverty and Inequality*. Washington, DC: World Bank. https://openknowledge.worldbank. org/handle/10986/11985Se.

Headey, Derek, and Olivier Ecker. 2012. "Improving the Measurement of Food Security." IFPRI Discussion Paper No. 01225, International Food Policy Research Institute, Washington, DC. https://www.ifpri.org/publication/improving -measurement-food-security.

Hill, Ruth, Emmanuel Skoufias, and Barry Maher. 2019. *The Chronology of a Disaster: A Review and Assessment of the Value of Acting Early on Household Welfare*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/31721.

Hillis, Susan D., H. Juliette Unwin, Yu Chen, Lucie Cluver, Lorraine Sherr, Philip S. Goldman, Oliver Ratmann, Christi A. Donnelly, Samir Bhatt, Andrés Villaveces, Alexander Butchart, Gretchen Bachman, Laura Rawlings, Phil Green, Charles A. Nelson III, and Seth Flaxman. 2021. "Under the Radar: Global Minimum Estimates for COVID-19-Associated Orphanhood and Deaths among Caregivers." *Lancet* 398 (10298): 391–402. https://www.thelancet.com /journals/lancet/article/PIIS0140-6736(21)01253-8/fulltext.

Honorati, Maddalena. 2019. "National Roll-Out of Albania Ndhima Ekonomike New Eligibility Rule: A Review of Results between January–July 2018." World Bank, Washington, DC.

Jensen, Robert T., and Nolan H. Miller. 2011. "Do Consumer Price Subsidies Really Improve Nutrition?" *Review of Economics and Statistics* 93 (4): 1205–23. https:// direct.mit.edu/rest/article/93/4/1205/57895/Do-Consumer-Price-Subsidies -Really-Improve.

Kakwani, Nanak, and Kalanidhi Subbarao. 2005. *Ageing and Poverty in Africa and the Role of Social Pensions*. Washington, DC: World Bank. https://openknowledge .worldbank.org/handle/10986/8535.

Komives, Kristin, Vivien Foster, Jonathan Halpern, and Quentin Wodon. 2005. *Water, Electricity, and the Poor: Who Benefits from Utility Subsidies?* Directions in Development. Washington, DC: World Bank. https://openknowledge .worldbank.org/handle/10986/6361.

Kosec, Katrina, and Cecilia Hyunjung Mo. 2017. "Aspirations and the Role of Social Protection: Evidence from a Natural Disaster in Rural Pakistan." *World Development* 97: 49–66. https://www.sciencedirect.com/science/article/pii /S0305750X17301092.

Lambert, Sylvie, Dominique van de Walle, and Paola Villar. 2017. "Marital Trajectories and Women's Wellbeing in Senegal." WIDER Working Paper Series 124, United Nations University World Institute for Development Economic Research, Helsinki, Finland. https://www.econstor.eu/handle/10419/189969.

Lindert, Kathy, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Nishikawa Chávez. 2020. *Sourcebook on the Foundations of Social Protection Delivery Systems*. Washington, DC: World Bank. http://hdl.handle.net/10986/34044.

Lloyd, Cynthia B., and Sonalde Desai. 1992. "Children's Living Arrangements in Developing Countries." *Population Research and Policy Review* 11 (3): 193–216. https://www.jstor.org/stable/40229805.

López-Calva, Luis F., and Eduardo Ortiz-Juarez. 2011. "A Vulnerability Approach to the Definition of the Middle Class." Policy Research Working Paper 5902, World Bank, Washington, DC. http://hdl.handle.net/10986/3669.

Mancini, Guilia, and Giovanni Vecchi. 2020. "On the Construction of a Welfare Indicator." For Inequality and Poverty Analysis, decision draft memo, Poverty and Inequality Global Practice, World Bank, Washington, DC.

Menezes-Filho, Naercio, Bruno K. Komatsu, and João Pedro Rosa. 2021. "Reducing Poverty and Inequality during the Coronavirus Outbreak: The Emergency Aid Transfers in Brazil." Policy Paper No. 54, INSPER Instituto de Ensino e Pesquisa, São Paulo, Brazil. https://www.insper.edu.br/wp-content/uploads/2021/02/Policy_Paper_54.pdf.

Moffitt, Robert A., Brian J. Phelan, and Anne E. Winkler. 2020. "Welfare Rules, Incentives, and Family Structure." *Journal of Human Resources* 55 (1): 1–42. http://jhr.uwpress.org/content/55/1/1.short.

MoLSA, World Bank, and UNICEF Armenia (Ministry of Labour and Social Affairs, World Bank, and United Nations Children's Emergency Fund Armenia). 2021. *Core Diagnostic of the Social Protection System in Armenia*. Washington, DC: World Bank and UNICEF.

Munoz Boudet, Ana Maria, Paola Buitrago, Benedicte Leroy De La Briere, David Newhouse, Eliana Rubiano Matulevich, Kinnon Scott, and Pablo Suarez-Becerra. 2018. "Gender Differences in Poverty and Household Composition through the Life-Cycle: A Global Perspective." Policy Research Working Paper 8360, World Bank, Washington, DC.

Newhouse, David, Pablo Suarez-Becerra, and Martin C. Evans. 2016. "New Estimates of Extreme Poverty for Children." Policy Research Working Paper 7845, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/25162.

Penglase, Jacob. 2021. "Consumption Inequality among Children: Evidence from Child Fostering in Malawi." *Economic Journal* 131 (634): 1000–25. https://doi.org/10.1093/ej/ueaa104.

Premand, Patrick, Pascale Schnitzer, and Dominique van de Walle. Forthcoming. "Intra-Household Inequality and the Selection of Cash Transfer Recipients: Evidence from Niger." World Bank, Washington, DC.

Ravallion, Martin. 1998. "Poverty Lines in Theory and Practice." Living Standards Measurement Study Working Paper 133, World Bank, Washington, DC. https://documents1.worldbank.org/curated/en/916871468766156239/pdf/multi-page.pdf.

Ravallion, Martin. 2015. "On Testing the Scale Sensitivity of Poverty Measures." *Economics Lett*ers 137: 88–90. https://www.sciencedirect.com/science/article/pii/S016517651500436X.

Rentschler, Jun, and Melda Salhab. 2020. "People in Harm's Way: Flood Exposure and Poverty in 189 Countries." Policy Research Working Paper 9447, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34655.

Rodriguez, Laura, and Matt Wai-Poi. 2020. "Fiscal Policy, Poverty and Inequality in Jordan: The Role of Taxes and Public Spending." Middle East and North Africa Poverty Practice, World Bank, Washington, DC.

Ruggeri Laderchi, Caterina. 2014. "Transitional Policies to Assist the Poor While Phasing out Inefficient Fossil Fuel Subsidies That Encourage Wasteful Consumption." Contribution by the World Bank to G20 Finance Ministers and Central Bank Governors. World Bank, Washington, DC. http://www.g20.utoronto.ca/2014/14%20Transitional%20Policies%20To%20Assist%20The%20Poor%20While%20Phasing%20Out%20Inefficient%20Fossil%20Fuel%20Subsidies.pdf.

Ruggeri Laderchi, Caterina, Nikola L. Spatafora, Sudhir Shetty, and Salman Zaidi. 2017. *Riding the Wave: An East Asian Miracle for the 21st Century*. World Bank East Asia and Pacific Regional Report. Washington, DC: World Bank Group.

Silwal, Ani Rudra, Solrun Engilbertsdottir, Jose Cuesta, David Newhouse, and David Stewart. 2020. "Global Estimate of Children in Monetary Poverty: An Update." Poverty and Equity Discussion Paper, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34704.

Smith, Lisa, Olivier Dupriez, and Nathalie Troubat. 2014. "Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys." Working Paper No. 008, International Household Survey Network, Western Michigan University, Kalamazoo, MI.

Trócaire. 2017. "Understanding Women's Lives In Polygamous Marriages: Exploring Community Perspectives In Sierra Leone & DRC." Trócaire and Irish Aid, County Kildare, Ireland. https://www.trocaire.org/sites/default/files/resources/policy/1-overview-report-polygamy_0.pdf.

United Nations. 1989. "Household Income and Expenditure Surveys: A Technical Study." National Household Survey Capability Programme, United Nations, New York.

Unwin, H. Juliette T., Susan Hillis, Lucie Cluver, Seth Flaxman, Philip Goldman, Alexander Butchart, Gretchen Bachman, Laura Rawlings, Christl A. Donnelly, Oliver Ratmann, Phil Green, Charles Nelson, Alexandra Blenkinsop, Samir Bhatt, Chris Desmond, Andrés Villaveces, Lorraine Sherr. Forthcoming. "Global, Regional, and National Minimum Estimates of Children Affected by COVID-19-Associated Orphanhood and Caregiver Death, by Age and Family Circumstance Up to Oct. 31, 2021: An Updated Modelling Study. *Lancet Child & Adolescent Health*. Published online February 24, 2022. https://www.doi.org/10.1016/S2352-4642(22)00005-0.

Vandermeersch, Celine. 1997. "Les Enfants Confies Au Senegal." Colloque Jeune Chercheurs, French Institute for Demographic Studies, Paris.

van de Walle, Dominique. 2013. "Lasting Welfare Effects of Widowhood in Mali." *World Development* 51 (C): 1–19. https://econpapers.repec.org/article/eeewdevel/v_3a51_3ay_3a2013_3ai_3ac_3ap_3a1-19.htm.

Velarde, Rashiel. 2018. "The Philippines' Targeting System for the Poor: Successes, Lessons and Ways Forward." Social Protection Note No. 16, World Bank and Australian Aid Policy, Washington, DC. http://hdl.handle.net/10986/30972.

Wilcox, Bradford, Joseph P. Price, and Angela Rachidi. 2016. "Marriage, Penalized: Does Social-Welfare Policy Affect Family Formation?" Institute for Family Studies, American Enterprise Institute, Washington, DC. https://ifstudies.org /wp-content/uploads/2016/10/IFS-HomeEconReport-2016-Final_Digital.pdf.

World Bank. 2003. *Measuring Living Standards: Household Consumption and Wealth Indices*. Washington, DC: World Bank.

World Bank. 2013. *Pakistan's Citizens Damage Compensation Program (CDCP)*. Washington, DC: World Bank. https://www.gfdrr.org/sites/default/files /Pakistan%20Citizens%20Damage%20Compensation%20Program%20 %28CDCP%29.pdf.

World Bank. 2016. *Jordan Economic Monitor, Fall 2016: Reviving a Slowing Economy* . Washington, DC: World Bank.

World Bank. 2018a. *Poverty and Shared Prosperity Report 2018: Piecing Together the Puzzle*. Washington, DC: World Bank. https://openknowledge.worldbank.org /bitstream/handle/10986/30418/9781464813306.pdf.

World Bank. 2018b. *Riding the Wave: An East Asian Miracle for the 21st Century*. World Bank East Asia and Pacific Regional Report. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/28878.

World Bank. 2019. *Aspiring Indonesia: Expanding the Middle Class*. Washington, DC: World Bank. http://hdl.handle.net/10986/33237.

World Bank. 2020a. "Brazil: Income Support for the Poor Affected by COVID-19 Project." World Bank, Washington, DC.

World Bank. 2020b. *India: Accelerating India's COVID-19 Social Protection Response Program*. Washington, DC: World Bank Group. https://hubs.worldbank.org /docs/imagebank/Pages/docProfile.aspx?nodeid=32015405.

World Bank. 2020c. *India: Second Accelerating India's COVID-19 Social Protection Response (PMGKY) Development Policy Operation*. Washington, DC: World Bank Group. https://imagebank2.worldbank.org/search/32602941.

World Bank. 2020d. "Morocco: Covid-19 Social Protection Emergency Response Project Appraisal." Document 4076, World Bank, Washington, DC. https:// imagebank2.worldbank.org/search/32602938.

World Bank. 2020e. "Philippines Beneficiary First Social Protection Project." Project Appraisal Document 3951, World Bank, Washington, DC. https://imagebank2 .worldbank.org/search/32417210.

World Bank. 2020f. *Poverty and Shared Prosperity Report 2020: Reversals of Fortune*. Washington, DC: World Bank. https://www.worldbank.org/en/publication /poverty-and-shared-prosperity.

World Bank. 2021a. "Pakistan: Crisis-Resilient Social Protection (CRISP) Program-for-Results Project." Washington, DC: World Bank Group. https://imagebank2 .worldbank.org/search/32901147.

World Bank. 2021b. "Support to the Development of Colombia's Integrated Social Registry." Report No. PAD4202, World Bank, Washington, DC.

4

Improving Targeting Outcomes through Attention to Delivery Systems

Margaret Grosh, Phillippe Leite, Emil Tesliuc,
Nina Rosas, and Priyanka Kanth

Introduction

Delivery systems are important for reducing errors of exclusion and inclusion, and for ensuring good implementation and dynamism. The importance of "implementation" in the context of targeting is suggested by Coady, Grosh, and Hoddinott (2004) and written about more explicitly by Devereaux et al. (2017) and Leite et al. (2017) and in a great deal of the program or country-specific case literature. Lindert et al. (2020) go far in codifying knowledge and improving a shared language around delivery systems. The volume also underscores the commonalities in delivery systems and their workings across many programs. Governments have made significant strides in this field in recent years, but a substantial need for improvement still exists, especially with respect to inclusion and dynamism.

This chapter focuses on delivery systems before discussing the choice of targeting methods, to emphasize the importance of implementation of different elements of the delivery chain for improving targeting performance, especially lowering errors of exclusion. No matter how aptly selected the targeting method is, it cannot deliver good outcomes without good implementation of each step of the delivery chain. Indeed, understanding how

crucial delivery systems are comes in part from literature on age-based social pensions or child allowances, which, despite the simplicity of their eligibility criteria, struggle with some of the same practical issues as programs with more complex eligibility criteria to get to the desired level of inclusion.

In 2001, Bolivia's universal social pension, for example, had an overall coverage rate of just 70 percent of the elderly and only 37 percent of the elderly in the poorest quintile, due to issues of lack of information, lack of identification (ID), and distance from pay points, which were more binding constraints for the poorer. The country made an effort to resolve these issues; as a result, only two years later, coverage was 79 percent overall and 58 percent among the poorest quintile. Overall coverage continued to improve to 90 percent and fully 90 percent of the poorest quintile, showing how sustained effort on delivery systems can improve outcomes (Muller 2016; Rofman and Oliveri 2012).

Nepal is also on a path to deliver its age-based programs, with more mature programs for social pensions for senior citizens and single women. Geographic coverage of the relatively newer child grant for children younger than five years, which was initiated in 2009, is being gradually expanded. A diagnostic report commissioned by the United Nations Children's Fund to review registration in these programs (Thinkthrough Consulting 2021) reports higher coverage of programs for senior citizens and single women, at 85 percent of the eligible, compared with the child grant, at 50 percent, as of the 2019 Multiple Indicator Cluster Surveys. The report recommends actions along the delivery chain congruent with those explained in this chapter—strengthening outreach and awareness among the population, especially those with language and remoteness barriers; lowering transaction costs; improving grievance redress; and improving institutional capacities through better staffing, training, digitization, and internal monitoring procedures.

Although the details differ, most social protection programs follow a common delivery chain, and people can be incorrectly included or excluded at various steps (figure 4.1). The phases in the delivery chain that are common to most programs include outreach, intake and registration, assessment of needs and conditions, eligibility and enrollment, payments of benefits and provision of services, and beneficiary operations management, including beneficiary exits. People can be wrongfully left out or brought in due to implementation failures at any step, from the definition of the intended population to the caseload of enrolled claimants, generating targeting errors, which then filter through to subsequent stages. Thus, distributional outcomes depend on the entire delivery chain; all of these steps matter, not simply a single point at eligibility determination.

Figure 4.1   Social Protection Delivery Chain
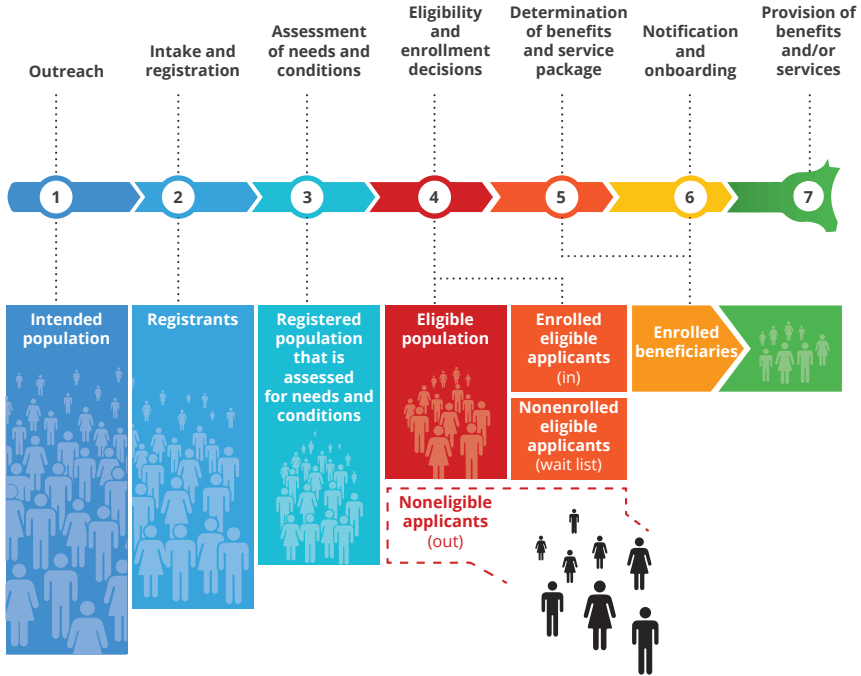
Source: Lindert et al. 2020.

People and institutions interact all along the delivery chain, and those interactions are facilitated by communications, information systems, and technology. Often the headline attention goes to information systems and technology. The information technology (IT) revolution can indeed be transformative for some aspects of delivery systems, but effective delivery systems also require sound rules (which the IT system may help to implement) and staffing to help make those rules and handle aspects of the job that need a human touch. The goal of social protection and its delivery systems is to make people's lives better. Thus, it is important to understand how people journey through the delivery chain and address problems that make that journey difficult. Some of the solutions can be addressed by IT and others cannot.

Good delivery systems are important for compliance with several of the principles of the human right to social security. This chapter identifies many aspects of implementation that support accessibility, dignity and autonomy, nondiscrimination and equality, inclusion of vulnerable groups, gender sensitivity, and transparency and accountability as they are understood in

the right to social security.[1] For example, providing physical accessibility is specifically called for in Comment 19 on the right to social security of the Economic and Social Council, and clearly it is helpful in reducing transaction costs and errors of exclusion. Likewise, providing materials and staffing for various languages as needed is important for nondiscrimination and dignity and the inclusion of vulnerable groups such as indigenous groups, ethnic minorities, and immigrants. Providing clear information on processes can help people know whether and how to apply or appeal, which will lead to high inclusion and be in keeping with the transparency and accountability standards of human rights. Ensuring that all processes are effectively accessible to women is in keeping with gender sensitive social protection—which may imply working to remedy gender gaps in documents for identification, providing female interviewers or intake officers, and ensuring that grievance-redress mechanisms pay attention to gendered power differences, which may discourage women from voicing their concerns or lodging complaints. Indeed, a great deal of the bad reputation of targeting with respect to human rights is earned through insufficient delivery systems rather than inherent in the process of eligibility determination. Human rights perspectives can be quite useful in spurring or guiding improvements in delivery systems.

Minimizing process-related targeting errors, especially errors of exclusion, requires significant political will, management attention, and administrative budget, which are factors that all too often have been scarce. One of the concerns about choosing to differentiate eligibility by welfare is the administrative costs implied, which, as chapter 2 shows, are not prohibitive. However, concerns about transaction costs and stigma leading to errors of exclusion, reduced program impacts, and loss of political support can all be made worse by poor delivery systems. This chapter suggests many ways in which improved delivery systems can ameliorate some of these costs of targeting.

The importance of delivery systems is underscored every time there is a crisis or disaster. The tasks to be done—outreach, assessment, payment, and monitoring—still need to be done but with more urgency. Emergency responses can be hampered or facilitated by how well developed the basic building blocks of the delivery system are—coverage of unique IDs and bank accounts; coverage, recency, and pertinence of information in a social registry or integrated information system; capacity of staff; and so forth. This means that most improvements made in ongoing programs and during stable times are likely to pay off doubly, for their base use and for their ability to support emergency response. In addition, crisis adaptations can be facilitated with advance planning.

This chapter unpacks aspects of the delivery chain that are important for targeting performance without rehashing the available evidence on

current practices. Many of the lessons provided in the chapter are a synopsis of those from Leite et al. (2017) and Lindert et al. (2020) and, for the crisis sections, from Beazley, Solórzano, and Barca (2019) and Bowen et al. (2020), although the illustrations from recent program practice aim to supplement those or emphasize their targeting aspects. Those fuller documents are considered companion volumes to this one, especially Lindert et al. (2020). The treatment here is brief and meant to provide enough of a framework to illustrate the need to delve deeply into the topic.

The chapter is organized as follows. The first section follows the delivery chain, highlighting at each step considerations that are important in ensuring that all who are in the group meant to be served actually are served, with dignity and at moderate administrative costs. The second section takes up some of the considerations needed to make delivery systems responsive to major shocks. The third section discusses some of the institutional issues that shape the client interface for the delivery of social assistance. The final section considers back-office issues in data management and data protection.

Fortifying Weak Links in the Delivery Chain to Reduce Errors of Exclusion and Inclusion

The first phase of the delivery chain, outreach, involves communication to inform the intended population about one or more social programs for which they may be eligible and the processes for registering. If registration processes serve just one social program, then the communications would focus on the main features of that program (objectives, eligibility criteria, rights, and responsibilities). If the registration processes are common across programs in a shared social registry, then the communications need to cover the main features of those diverse programs or at least the notion that the social registry helps connect people to programs. The effect of communication on targeting begins at the outreach phase, but it elicits (or fails to elicit) actions from the intended population at every phase and thus must be embedded into each phase rather than being viewed as separate from the rest of program implementation.

Success in the communications phase would be that people can make a well-informed decision about whether to apply for a program and what it takes to do so. Success on these two fronts can lower transaction costs for people and administrative costs for programs at later steps in the delivery chain. People who are not eligible for clear reasons (wrong age, location, or clearly do not meet eligibility criteria) may understand that and therefore not waste time and hope in applying. Those who are likely to be eligible would see a benefit in doing so. They would understand how to apply or how to obtain further details in simple ways so that they need not suffer

extra transaction costs by journeying and queueing more times than needed at different offices. Moreover, well-informed decisions by potential clients will reduce the number of unsuccessful applications social protection program officers would need to handle and the administrative costs involved.

Communicating proactively in ways that are sensitive to the clients' needs and barriers is vital to success throughout the delivery chain. Comprehensive communication strategies include multiple channels and are tailored to the target population. For a busy urban population in a high-income setting, websites and call centers may help provide information at flexible times and quickly, although even in this setting, the digital divide and lack of baseline knowledge must be overcome. Broadcast channels may include radio, television, social media, or community theater; they may include posters and pamphlets in well-selected locations, such as markets, community centers, places of recreation or houses of worship, and shared water points or bus stops; and they may work through trusted agents in contact with the intended population, such as community leaders, religious leaders, teachers or health workers, or peers such as fellow mothers or youth leaders. It may be helpful to work with advocacy or nongovernmental groups that provide services to groups commonly facing barriers to inclusion, such as people living with disabilities; people who are lesbian, gay, bisexual, transgender, queer (or questioning), and others; homeless people; or others. The messages and media should be well tailored to the audience's language and level of literacy and be phrased in ways to encourage take-up and discourage stigma. They should be inclusive of those with vision or hearing impairments. Simplicity and clarity are helpful for those who are stressed by poverty or crisis or have limited experience of agency with respect to bureaucracy. Communications should be repeated or reinforced at judicious intervals. A person might miss one message and it may take more than one seeing/hearing of a message to persuade someone that it is worth taking action to apply for a program. Timing could be important—with messages repeated at times of heightened need, such as the beginning of the lean season or school year, and many countries have boosted outreach in response to the COVID-19 crisis.

Active outreach using a variety of agents can promote potential inclusion of marginalized groups who may be unaware of the processes for inclusion in social programs. This typically requires special efforts to find potentially eligible people who are likely to be left out of more mainstream efforts, in their homes, places of work, or places of leisure. Active outreach strategies rely on the deployment of teams at the local level to focus on identifying those who may not be able to enroll in programs in city centers and in particular geographic areas that are remote, hard to reach, or comprised of a population that is consistently marginalized. Brazil's Busca Ativa outreach program, which is described in chapter 2, is well known. Pakistan

added street theater and outreach by "mother leaders" who were active claimants in the income support program (box 4.1). The Philippines' Listahanan social registry is known for its positive branding efforts (Lindert 2017; Velarde 2018). World Bank and DFAT (2020) cites traditional problems but improving practice in outreach in East and South Asia.

BOX 4.1 *(continued)*

response in 2020, the government launched the Ehsaas Emergency Cash Program, which aimed to provide PRe 12,000 (US$75 at the time; US$68 in 2022) per family to 15 million families, including the existing 4.6 million families under the Ehsaas Kafaalat. The government employed mobile phone messaging services and other new communication tools to collect and provide eligibility information from and to potential beneficiaries and reduce physical touch points for intake, registration, and payments. Physical delivery points were managed through proper social distancing measures.

Source: Prepared by the National Social Protection Program for Results (P158643) Task Team, World Bank.

As with outreach, intake and registration processes should consider language, mobility, and cultural (ethnic, religious, or other) considerations. People of diverse backgrounds and especially those in the population that a given program is meant to serve should find it straightforward and comfortable to apply for benefits that might be due to them. Forms and interfaces may need to be built in multiple languages, and staffing should accommodate the principal languages spoken in each area of the country, with some provision for interpretation of less commonly spoken languages. Ideally, program staff will reflect the cultural background of clients to reduce any issues of discrimination, ill-treatment, or stigma. Mexico actively recruited and trained indigenous people, mostly women, from program areas for the PROGRESA-Oportunidades-PROSPERA program.[2] In Brazil, some social assistance centers recruit quilombolas[3] as staff to avoid barriers among quilombola community clients. Bulgaria makes an effort to use Roma staff to facilitate access by Roma clients. In the Republic of Congo, the intake forms and interviews can be done in Lingala as well as French. In Mali, they may be done in Bambara or French. Myanmar uses locally recruited facilitators in Shan state in its new maternal cash transfer program.

Intake and registration may be done "on demand" in an office or through online applications and/or in "administrator-driven" or "census sweep"–style field operations, each carrying risks of errors of exclusion that must be mitigated.

- On-demand application processes are normally open continuously. They may use a combination of physical office, call-in, or virtual application tools. Continuously open application processes are vital to the inclusion of newly formed families, families new to the location, or families newly

in need. However these processes rely on the initiative of the applicant to apply and are thus at risk of errors of exclusion. Thus, outreach is vital so the potential applicant would know how to apply, and transaction costs to applicants must be low. Using interoperable systems to refresh information regularly from other government databases is a common element of on-demand application and recertification processes. On-demand processes also lend themselves to differentiated periods for recertification, by program or client characteristic.

- Administrator-led processes go to the field periodically, usually in a census sweep, and thus may help reach those who would not have known to apply or found it easy to do so. These administrator-led processes are mostly associated with programs that use a proxy means test (PMT) or community-based targeting or a combination to determine eligibility, although they are occasionally also used at the initiation of demographically targeted programs. These mass field mobilizations sometimes combine in a single operation several steps of the delivery chain—even from outreach to onboarding. But they will contain inherent errors of exclusion during the periods between mobilizations if households do not have a way to request registration. This will especially affect newly formed households, households new to the area, or those that have suffered a decline in welfare since the prior targeting exercise. Because of this, even programs with periodic administrator-led processes should also have on-demand processes and outreach, a feature that is all too often missing.

An informal survey of the state of social registries showed that just over half of the countries, mostly upper-middle-income countries and those in Eastern Europe and Central Asia, have dynamic registries. With some notable exceptions (for example, Argentina, Brazil, Chile, and Uruguay, and most recently Colombia, which is in the process of converting to a dynamic system), most of the countries in Latin America still have static registries despite many of them having been established many years ago. Africa's registries too are mostly still static, although the registries are younger and in lower income countries, so this may be a natural first step.

Intake and registration involve the process of collecting self-reported information through an application form or questionnaire and documentation to register the intended population for consideration of potential eligibility for social programs. The form should be user-friendly: not too long in terms of the number of questions or time taken to administer, and easy to comprehend and navigate. To reduce transaction costs to clients, this process should avoid gathering (again) complex information that can be drawn from other databases through data matching. The basic components of information to collect fall into the following categories:

- *Information on the social assistance unit.* This contains at least the name and pertinent ID number (or age, place of birth, or other identifiers) of at least the individual(s) who comprise the assistance unit. Even for programs focused on a specific individual, it is often valuable to gather information on all members of the assistance unit or household and how they are related, for example, for cross-referrals to other programs, to build a comprehensive measure of welfare, or to link the target individual (child) with the legally responsible adult(s).
- *Contact information.* This information includes the applicant's mobile phone, email, physical address, or georeferenced coordinates. At a minimum, the program administrators need a way to communicate with the applicant. Location information may also serve in assessing risks, needs, or welfare.
- Sometimes to minimize the number of contacts needed with the client to finish enrollment, the application form also includes *information pertinent to payments*. Strictly speaking, this information is not necessary at intake, and its collection may be deferred until enrollment. Information in the first three categories may be sufficient to determine eligibility for programs based only on age or location.
- For programs where eligibility is determined by socioeconomic status, such as by means, hybrid, or proxy means testing, intake and registration should also include *sociodemographic and socioeconomic variables* that allow program administrators to obtain the actual welfare or estimate it based on the characteristics of the household members (chapter 6 provides details on how to measure or estimate welfare).[4]
- Given the role of social programs in preventing/mitigating and responding to shocks, some administrators are beginning to collect additional information at intake, such as *exposure to natural hazards*.

The required documentation should be kept to a minimum to lower transaction costs and errors of exclusion. Although supporting documents can be required on any of the above topics, it is important to avoid requiring excessive documentation as each item may generate transaction costs to the applicant or be impossible for them to achieve, thereby generating exclusion. When data exchange with other government databases can be set up, this can lower the transaction costs to applicants and potential errors of exclusion. For example, it is preferable not to ask individuals to supply pay stubs, tax statements, titles to government-registered assets, or benefits from other government social programs, but to draw on data matching to access such information where it is required for eligibility determination. Tesliuc et al. (2014), for example, report that in Uzbekistan, about 25 percent of the poorest quintile report the number and complexity of documents and forms to be filled out as one of the reasons for not applying for benefits. At the opposite end is Albania, where applicants must submit only two

documents in addition to filling out a declaration with the required information, which is then subject to verification through cross-checks with other institutions and agencies.

All social assistance programs require applicants to declare and prove their identity. This could be simple or not, depending on the country's identity ecosystem and degree of integration with the social assistance program. The drive toward higher coverage of foundational ID systems, and especially electronic identification (eID), hugely facilitates some of the goals and processes of social assistance. However, even in countries where the coverage of foundational IDs is overall high enough to be used in normal enrollment processes, there is a risk of important exclusions if registration processes do not provide alternatives for those who do not have the first-choice document. Court cases in several countries have upheld the right to services without ID, with India being notable but not alone among them.[5] Since in general those who are not included in the ID system are those who may be at high risk of other exclusions that may be associated with poverty or risk, it is important for distributional outcomes.[6]

The availability of foundational ID registries or their electronic variants reduces both the transaction costs of social assistance applicants and exclusion and inclusion errors. Foundational ID systems[7] provide identity credentials for all the resident population, irrespective of citizenship or immigration status. Examples of ID credentials include birth certificates provided by the civil registry and ID cards provided by a population or ID registry. While traditionally these credentials are in paper form, recently many countries have migrated from paper to electronic registries, from simple paper credentials to smart ID cards with biometric or other secured elements, or even to online biometric authentication systems. According to the Global ID4D Dataset 2018,[8] 96 percent of high- and middle-income countries have at least one form of digitized ID registry, but only 70 percent of the low-income countries have such a program. When such e-registries are available and interoperable with the social program, there are several advantages. First, the e-registries deduplicate their records to ensure that one individual has only one ID record. Second, a social program could check online if the ID provided by applicants is genuine or not, and if it is, the ID system will guarantee that it is unique. Deduplicated IDs can reduce the risk of fraud through duplicate or ghost beneficiaries and lower errors of inclusion. Third, the use of a foundational ID system implies that the ID used for the program can also be used to comply with know-your-customer standards in the finance sector and facilitate digital payments, which can lower administrative and transaction costs and increase security.

Countries without foundational ID systems (other than a paper-based civil registry) have created a variety of functional ID systems to manage identification, authentication, and authorization for specific sectors or

use-cases, such as social protection. For example, in the United States, identity is proved through an individual's social security number and driver's license. In most low-income countries and some middle-income countries, this would result in incomplete coverage of the population with ID credentials. In these countries, social programs use more limited functional documents without the potential benefits of data matching and know-your-customer verification.[9]

Incomplete coverage of the population with ID documents or other credentials may be a barrier to participation for some in important target groups, such as women and people who are poorer, more remote, and less literate (see, for example, Gelb and Metz 2018; World Bank 2018b). Data from the Global ID4D Dataset 2018 indicate that an estimated 1 billion people do not have an official proof of identity; nearly 50 percent of those are in Africa; and 40 percent of people in low-income countries lack IDs (World Bank 2018a). The data suggest a gender gap in low-income countries, where close to 44 percent of women lack IDs, compared with 28 percent of men. The importance of remedying ID issues is reflected in the Sustainable Development Goal target on the topic and has sparked initiatives such as the World Bank's ID4D program to improve coverage of IDs throughout the world and especially in developing countries. In various countries, efforts to improve the coverage of IDs work in tandem with social assistance programs that aim to reach the people who are least likely to have IDs. In Pakistan, the Benazir Income Support Program and National Database and Registration Authority cooperated extensively to provide the Citizen National Identity Card to women, with the number of registered women nearly doubling from 21 million to 40 million between 2009 and 2012.[10] Following the introduction of a child grant in the poorest districts, birth registration increased from 40 to 90 percent (Amjad, Irfan, and Arif 2015). Yet, in many countries, the ID issue is not yet solved, particularly for the groups that are of most concern for poverty targeting and other social assistance. Care must be taken to work to remove these barriers to foundational IDs and/or provide workarounds, such as accepting various forms of functional IDs or providing program-specific functional IDs.

The need to prove (legal) residency or provide an address can be another important barrier to program participation, and it must be considered critically. Social programs often request proof of residency. This is especially the case for programs implemented by lower levels of government where financing is local or depends on a rationed allocation of budget or program slots from a higher level of administration. This may be done to prevent double-dipping, where a person claims benefits from two jurisdictions, or to conserve resources to match local allocations. Some programs may request an address as part of establishing ID or a way to contact clients. Moreover, PMT formulae usually consider aspects of the applicant's

dwelling as part of the needs assessment. These requirements each serve a seemingly useful administrative function, but they can also create barriers of exclusion for families that move from one jurisdiction to another or for the homeless. In China, for example, people without local household registration, *hukou*, are not entitled to benefits under the *dibao* social assistance program, and *dibao* benefits are not portable across jurisdictions, rendering migrants excluded by design.[11] In India, none of the flagship national social assistance programs have traditionally been open to cross-state migrants, although as part of the COVID-19 policy response, India plans to make benefits in the Targeted Public Distribution System portable (World Bank 2020a). In Albania, claimants who do not have a formal residence or domicile in the locality where they reside de facto are required to apply in the locality of formal residence or submit documents issued from those localities. Thus, internal migrants are not excluded by design, but the requirements may represent barriers for the poor in accessing the program (Tesliuc et al. 2014). In Colombia, a court case in 2016 prompted the National Planning Department that operates the PMT-based social registry to work with municipalities to develop a registry of the homeless and find ways to associate them to programs (DNP 2016). Although a mailing address has been the traditional way of providing a point of contact, there are a range of options for those with insecure housing, such as allowing claimants to use the address of a nongovernmental organization or social services provider or a post office box, cell phone, or email address. It may be easier for field workers to use a set of Global Positioning System (GPS) coordinates, since addressing is difficult in rural areas and informal urban settlements. This alternative may provide advantages in merging with geo-tagged data sets that are pertinent to disaster hazard or maps of delivery of allied social services.

Where the assistance unit is the family or household rather than the individual, there will need to be some register of the membership of the social assistance unit. Some countries have family certificates or registers already established for purposes outside social protection,[12] which can facilitate the use of "family" as the social assistance unit and aid in various parts of delivery systems, but many countries do not. Most family/household-level registries are formed as part of the application process for a social program or registration in a social registry. Family and or household composition is self-declared as most countries do not have family or household registries. When civil registration is fully automated and IDs are widespread, civil registration allows cross-verification of the declared information through the interoperability of systems. However, this process misses undeclared cases, such as children who left home and live with partners without declaring a change in residence or the fact that they entered some form of union. Based on the self-declared information,

a unique number is generated for the family or household with a listing of the IDs for each member (be it foundational or functional). In some cases, the individual number of the head of household can be used as the house- hold number, although this is less desirable because it impedes mapping together information from different programs with different assistance units or consolidating a picture of the income and assets of all family members. Because membership in the social assistance unit is dynamic, there should be a way (and a requirement) for families to register such changes. In Organisation for Economic Co-operation and Development (OECD) countries that quantify the sources of error and fraud in social protection programs, misreporting of the composition of the assistance unit and the identity of its members is the second largest factor, after non- declaration nondeclaration or under-declaration of income and assets. This was the case in the United Kingdom in 2017–18, where misreporting of household composition or living together with another earner were the second reason for the estimated error and fraud rate for four means-tested programs; the first cause was misreporting of income or assets.[13]

Social registries shared across multiple programs are a commonly used tool, although many are still nascent. Barca and Chirchir (2014, 2017) and Leite et al. (2017) reveal considerable diversity in the typologies and trajec- tories of these systems with respect to their (1) institutional arrangements (central and local); (2) use as inclusion systems (single or multiprogram use, static or dynamic intake and registration); (3) structure as information systems (structure of data management and degree and use of interopera- bility with other systems); and (4) coverage, which ranged from 75 percent of households or more in a quarter of the cases to less than 10 percent of households in another quarter of the cases (box 4.2).

Using a shared social registry as the entry point for multiple social pro- grams has benefits and risks for targeting outcomes. By providing a shared format, it harmonizes the information collected and can add coherence across social policy. By serving as a common portal, it can lower the costs of application as a household may have to apply only once to receive multiple sets of benefits, or at least receive cross-referrals that improve their knowledge of programs from which they may benefit. Similarly, shared registries may lower governments' total administrative efforts toward out- reach, intake, and registration. By uniting the efforts across various programs, governments may be able to amass resources and gravitas to do the work well. However, concentrating provision also concentrates risks as any failure in outreach or process affects not just a single program but many. The heavier is the use of the registry in social policy, the more important it be dynamic, inclusive, and accurate, and its data well pro- tected. Assessment of the needs and conditions of applicants is the part of the whole delivery chain that is most clearly associated with "targeting" or

BOX 4.2

How Big Should a Social Registry Be?

Social registries or interoperable social information systems range from covering just a few percent of the population to nearly all (figure B4.2.1). A social registry that is built to support just one or a few very narrowly targeted programs may cover just a few percent of the population. But the increasing trend is for social registries to be used to support a wide range of social programs, and some of these may be more broadly targeted, with some even reaching high up the welfare distribution. Cash transfers may be targeted at the poor—covering only 10 percent or so of the population in countries with very low poverty or very low budgets. But child allowances or social pensions may strive to include all members of the age group. Subsidies for health

Figure B4.2.1   Coverage of Social Registries
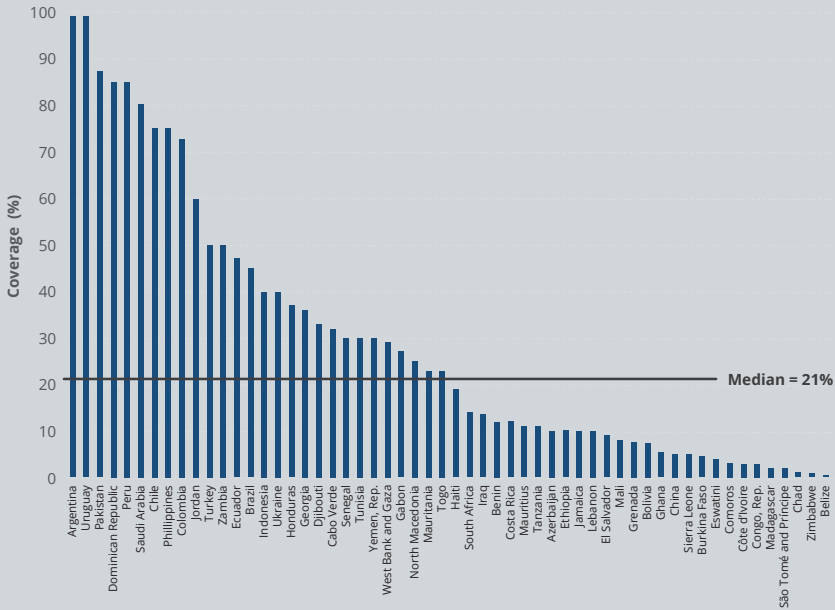
Source: Delivery Systems Global Solutions Group, Social Protection and Jobs Global Practice, World Bank.

BOX 4.2 *(continued)*

insurance may reach half or more of the population. Lifeline or "social" tariffs for utilities may be narrow or broadly targeted. Compensation in cash or in social tariffs instituted as part of the reform of food or energy prices can start as broadly as universal, with some sort of self- or affluence-targeting reducing participation at the top.

Social registries or program-specific information systems that include large shares of the population also offer flexibility for shock-responsive programming. With information already gathered on base welfare, contact points, and possibly accounts for payments, it is easier for the government to issue payments in response to shocks, with or without any additional needs assessment based on the household's welfare following the shock. Countries with large social registries have used this capacity in times of natural disaster: the Philippines is a prime example, and Kenya and Mauritania are countries where the dimension of the social registry in areas prone to drought was planned around the ability to respond when needed (see Bowen et al. 2020). With the COVID-19 crisis, quite a significant number of countries issued emergency payments to people contained in the social registry; where the registry was more inclusive, so too could be the response, covering not just the already poor but the many more affected by the various degrees of stay-at-home orders, closures, and quarantines imposed. Chile's temporary family emergency income support, for example, was targeted principally to families with only informal sector incomes, below the 90th centile of the welfare distribution and the 80th centile of the emergency index, and with reduced payments to those with some formal sector income. Payouts were automatic for those enrolled in several of the ongoing programs and required online applications for others (https://www.ingresodeemergencia.cl/faq).

Of course, there are costs to bear in having large social registries. The most obvious cost is that it takes resources to collect and update data. Thus, it is not natural for social registries to be much larger than the size of the largest ongoing or common emergency response program the registry is meant to support, although the registry may grow as the social protection system grows. A very large social registry may also be problematic politically, or it may require very good communications, to register many households who may be disappointed not to receive any support immediately. When Sierra Leone initiated its Social Protection Registry for Integrated National Targeting in 2014, it covered only a slightly larger number of households (approximately 20 percent) than could be covered by the main social assistance programs, to avoid raising expectations of a more expansive registry.

continued next page

In Mauritania, the provision for the emergency response program involved preregistering 50,000 households in drought-prone communes. In Kenya, the saturated coverage for the Hunger Safety Net registry is in only the four most drought-prone counties, and periodic payouts have been made using the information.

differentiating eligibility and benefits. Considerations around the selection of a targeting method and the details of its design are discussed in chapters 5 and 6. The information gathered in prior steps and any complementary data brought in from other government databases is compared with the eligibility criteria for one or several programs. The main output from this phase of the delivery chain is the preparation of a list that informs program administrators about the potential eligibility for specific programs or the mix of benefits and services that may be awarded. Another important output of this phase is the preparation of profiling reports on applicants, which allow measurement of the potential demand for social protection programs and determine the characteristics of the applicant population. As such, assessment of needs helps in planning, budgeting, and coordinating programs. Moreover, statistical tools, such as predictive analytics and data integration and analytics, can be used to predict the risks faced by different households and be helpful in disaster risk management.

Following assessment of needs and conditions, the next steps in the delivery chain are the conclusive steps of eligibility determination and enrollment decisions, notification, and onboarding. As presented in figure 4.1, exclusion errors are still possible if process failures occur at these stages. After the decision is made about eligibility to participate in a program, individuals, families, or households are classified as beneficiaries, wait-listed, or ineligible. This requires clear and proactive procedures for notification of eligibility as people need to be informed about their status, and miscommunication between administrations and people can undermine a program's credibility and transparency. Any applicant who does not understand what further steps they need to take to complete enrollment (or, if needed, to file a grievance) may yet lapse into errors of exclusion.

For those who are eligible, notification should include the next steps and procedures for official program enrollment. Notifications for this group should indicate the decision; what the beneficiary will receive; when, where, and how they will receive it; their rights and responsibilities; contact points and information; and next steps. At this enrollment point, eligible

beneficiaries need to have a more detailed and operational understanding of how the program works; who to contact; where and how to get benefits and services (payment points and service providers); payment and service provision schedules; the timing and location of any monitoring meetings; their rights, roles, and responsibilities; where and how to file grievances; and so forth. In this phase, some programs may require presentation of documents such as a photo (if a program-specific ID is created), cell phone number, bank account or e-wallet information, any relevant consent forms signed by the beneficiaries (or designated recipient), school enrollment form, medical certificates, vaccination cards, and so forth. There is a risk that applicants who are notified as being eligible may not be able to provide all the paperwork or that with inadequate notification, eligible people will lose benefits because they have improperly understood their responsibilities and fail to comply with them, thus leading to targeting error.

Take-up issues in the enrollment stage are like those seen during the intake and registration phase and require a strong understanding of the target group. Minimizing exclusion errors during notification requires a careful understanding of the capacity, effective communication channels, context, and constraints faced by the population of interest. For instance, designing literacy-appropriate notifications and enrollment materials is critical in most settings. This includes information on the mechanism to deliver the payments or benefits, which may be beyond what the potential beneficiary is accustomed to and can lead to issues downstream. Distance and associated costs may also be a constraint for potential beneficiaries.[14] Whether the notification process is public or private can also be important contextually. In some cases, there may be stigma associated with any type of public notification, while in others, a community meeting or other form of public notification is essential for the program's credibility or transparency.

Waiting lists bear important consequences for errors of exclusion. Anyone who is wait listed meets the eligibility criteria but is excluded from the program due to lack of budget space, which is a clear and direct error of exclusion, a situation that reveals clearly when there is a disconnect between eligibility criteria and budget that requires a policy solution. Moreover, when waiting lists are substantial and/or waiting times are long, they may have second-round effects. Individuals will come to understand that applications may be unsuccessful, so they may not bother to apply. Governments or individual staff workers may become less assiduous in outreach efforts since the program is already full, so that hidden errors of exclusion may occur. Recurring wait lists indicate a problem that needs fixing at the policy/budget level. Until that is done, not even allowing a wait list for a budget-rationed program can be worse. The wait list pressures the government for appropriations and allows quick response when they are forthcoming. Brazil's experience with wait lists for Bolsa Familia is a

case in point. From 2011 to 2014, Bolsa Familia carried out its iconic Busca Ativa program to reduce errors of exclusion among the homeless, riverine, and ethnic minorities. Then the economic downturn began, and gross domestic product growth decelerated by 10 percentage points and turned negative in 2015 and 2016. Extreme poverty doubled. The government stepped up cross-checking data sources and recertifications to ensure that ineligible claimants were not crowding out eligible ones, which, together with an eroding value of the eligibility threshold and the benefit fixed in nominal terms, allowed the program to keep a zero wait list even with a flat number of households enrolled in 2017 and 2018. Eventually a wait list began to build in 2019, reaching 1.2 million families just prior to the COVID-19 emergency declaration. As part of the emergency response to COVID-19, the government inaugurated a temporary emergency program of larger coverage and benefit than the regular Bolsa Familia parameters and increased funding of the basic Bolsa Familia program to take up after the temporary program, enough to cover the additional 1.2 million families on the wait list (World Bank 2020b).

Grievance and redress mechanisms (GRMs) can help to reduce targeting errors as well, especially errors of exclusion. By giving people the capacity to provide feedback to program administrators, a GRM provides beneficiaries and the general public a voice in the program's administration and performance management. The first role of the GRM is to correct applicant- or claimant-specific mistakes. People can check that the information used in determining their eligibility or payment was correct and get it corrected if there was an error. Of course for this to have teeth, there needs to be budget to allocate space to cases found eligible upon the handling of a grievance. In some programs, the GRM may provide a venue to request a judgment-based exception to the basic eligibility procedures. GRMs may also serve as a mechanism that provides an alert about systemic problems so that actions can be taken to reduce them. If complaints are consistently filed about certain parts of the process, that may signal the need for improved information, streamlined processes, or higher processing capacity. If complaints are filed about the competence of or discrimination or abuse by specific frontline workers, supervisors can intervene with (re-)training, sanctions for misperforming staff, or restaffing to prevent future instances, just as they work to rectify damage done in the specific case.

Similarly, a subsystem for error and fraud[15] control can help reduce mis-compliance errors that could lead to both inclusion and exclusion errors. Social protection programs channel a large amount of public resources to potentially millions of beneficiaries, with complex eligibility and recertification rules. Amid these myriad transactions, it is impossible to operate a program that it is completely free of error and fraud. In five OECD countries reviewed by the United Kingdom National Audit Office (2006),

this fraction varied between 2 and 5 percent of the total social protection spending. Information from developing countries is scarcer. Romania carried out benefit inspections of six large and error and fraud risk–prone social assistance programs during 2011–13 and found rates of irregularities between 8 and 20 percent. The rate of error and fraud varied by type of program, being higher in programs with more complex eligibility criteria and/or reassessment requirements, such as those that vary eligibility and benefit levels across the population. For means-tested programs, the rate of error and fraud was about 10 percent.

For the error and fraud risk–prone programs, on top of the existing and likely fragmented measures to reduce error and fraud, the social protection ministry or program administration should develop a comprehensive, end-to-end system. Generally, such a system would comprise two parts: an administrative unit that performs data analytics to detect suspicious cases, and a unit that uses this information to investigate and correct the associated over- or underpayments. The volume of such referrals can be larger than the human capacity to handle them. In such cases, a triage is typically done, with checks or inspections focused on the cases with the higher potential losses times the ability to correct them. To reduce error and fraud, governments and/or program administrators may focus efforts on high-budget, risk-prone programs such as income replacement programs or proxy- or means-tested benefits.

A decade ago, such error and fraud systems were only present in OECD countries. They have recently spread more and more into middle-income countries. A typical example of detection and correction is exemplified by the data-driven fraud detection system in France (OECD 2020). Since 2012, the Family Allowance Administration (Caisse d'allocations familiales) uses data mining and predictive modeling to determine which beneficiaries may be at risk of committing fraud, by identifying cases with similar characteristics to those already identified as fraudulent. In addition, the Family Allowance Administration checks the validity of administrative documents with the issuers (banks, internet and telephone access providers, utility companies, and so forth), mostly by automated exchange of information. For the applications with the highest risk, the Family Allowance Administration deploys additional verifications, which could go up to sending certified inspectors to the homes of claimants to conduct inspections and face-to-face interviews to determine the veracity of their claim. In Romania in February 2021, the Social Inspection Unit of the National Social Assistance Agency carried out a joint review with the Employment Agency to assess the effectiveness of the work conditionality for low-income households (ANPIS/ANPOFM 2021). The review used data cross-checks for the entire caseload of guaranteed minimum income beneficiaries, which were subsequently used to identify high-risk cases for in-person follow-up. In

Moldova, the Social Protection Ministry has a process for inspection based on a risk-profiling system for detecting errors and fraud, which includes home visits for a sample of households. This risk-profiling system is based on a statistical algorithm that flags cases that have a high likelihood of fraud. Since its introduction in 2009, the program has maintained low levels of errors of inclusion, with 80–90 percent of benefits accruing to households in the poorest quintile (World Bank 2018c). Other examples of error and fraud systems and their use are provided in Lindert et al. (2020) and Van Stolk and Tesliuc (2010).

For programs without fixed time limits for the duration of the social assistance unit's benefits, reassessment of eligibility (often called recertification) is logically required from time to time and important to maintain targeting performance over time. As illustrated in chapter 3, both the passage of time—with possible changes in households' income, demographics, and surrounding services—and positive effects of the programs themselves may change the household's welfare and thus eligibility status. To neglect this would certainly lead to errors of inclusion as some families prosper. If there are no new intakes, there will be errors of exclusion as other families fall into hardship, are formed, or move into the location. Even if there are some provisions for new intake but budgets are rationed, the continuing enrollment of families that were initially needy but no longer are will crowd out other now needier families.

As with the original registration processes, recertification may be done continuously or via periodic administrator-led processes, with powerful implications for logistics and targeting outcomes. Continuous recertification processes are more commonly used for on-demand programs and those that rely on interoperability with other government data systems rather than field work or home visits to gather information on welfare. Periodic survey sweeps or community-based targeting exercises are used in many countries as they establish large-scale programs and build social registries. This modality may endure for long periods, although a few of the early pioneers of survey sweep–based registries are finally moving toward more continuous processes.

Having a continuous process for recertification allows administrators to ensure that there are no periods of special friction when large groups of people are exited or a large wait list of eligible but unserved potential clients are excluded. With continuous modalities for recertification, it is logistically easy to recertify clients at different periodicities—for example, recertifying those whose employment and earnings might change more often than others, thus recertifying eligibility for young urban workers more often than for the elderly or people living with severe permanent disabilities. Frequent or differentiated periods for recertification are easiest when eligibility determination can draw heavily on administrative records that are updated

fairly often and automatically by other agencies and commonly seen means-tested guaranteed minimum income programs. Because people entered at different times, they come due for their periodic recertification in a continuous way. Handling their recertification requires a consistent level of staffing and should result in a fairly smooth flow of households in and out of the program each period, varying principally with periods of prosperity or recession.

Recertification by census sweep has the advantage of some economies of scale (for outreach and travel times) as the teams sweep through an area. However, this approach carries disadvantages. First, the big wave produces a spike in administrative costs, along with a need to mount, recruit, and train for a large field operation almost from scratch each time. Second, such big waves are done infrequently, contributing to exclusion and inclusion errors between the waves. It is not possible to customize for the different clienteles, and in the now nearly ubiquitous case that the social registry supports more than one social program, recertification is on the same schedule for all programs, irrespective of how that fits with the logic of the program. Third, the sweeps produce a large number of people to be moved off the roster at once, which may generate more political waves than a more routine process would. The greater is the number of people who might lose eligibility through a periodic recertification process, the greater will be the need for excellent communication and grievance redress mechanisms. It will also be more important that the cycle is regularly implemented and conducted during a low stress part of the political cycle, perhaps at the midpoint between elections, to minimize accusations of punishing or bribing voters. (Medellin et al. [2015] provide a summary of such issues for Latin American conditional cash transfer programs.)

Cost and capacity are obviously important in deciding how often to recertify and need to be balanced with the potential changes in targeting outcomes that would ensue in various cases. As chapter 2 discusses, the administrative costs of targeting have been kept manageable, in part by using fairly lengthy periods for recertification when it is done by field sweeps. However, chapter 3 notes that there are also significant losses in accuracy that come from this tactic. It is important to seek a balance. Each country (or program) should make some estimates for its own context and parameters, but as a rule of thumb, recertification periods should be no longer than every two or three years. More frequent recertification would make sense depending on the program and purpose, especially for programs that rely largely on data matching and interoperability. For programs or social registries that depend on new client contact for recertification, periods of more than five years would likely result in significant errors of inclusion and exclusion; therefore, investment in recertification every two or three years would likely pay off.

Regular assessments of the whole or specific parts of the delivery system through process evaluation, audits, and spot checks are important to inform program administrators about the strengths and weaknesses of the delivery chain. Identifying program implementation bottlenecks can help correct course to prevent systemic bias and challenges that affect targeting outcomes. Box 4.3 provides an illustration of how Mali used

BOX 4.3

Monitoring to Improve Performance in Mali's Jigisemejiri Program

Mali's Jigisemejiri program was launched in 2013 in an environment characterized by a paucity of administrative data.[a] The program offers a good example of how to develop a rich data system by collecting and analyzing information from program beneficiaries and other stakeholders, opening communication channels for data providers to visualize and check their information, and investing in audit or quality control activities that ultimately increase the accuracy, relevance, and use of the data. Although other databases were scarce and interoperability with them was limited, the program administrators at the Unité Technique de Gestion Filets Sociaux made use of field evidence based on a series of activities that improved the program's data system.

Several audits and data quality control activities were embedded in the program design by including spot checks to review the full implementation process. This included intake, registration, and targeting functions (UTGFS 2014b, 2014c, 2014d); quarterly reports on coverage, payments, and grievances[b] as well as profiling of the beneficiary population and assessments of grievances using the program management information system[c]; and posttransfer assessment (UTGFS 2014a) based on a random sample of about 800 beneficiaries to be selected from among the first 5,000 beneficiaries, to measure the degree of satisfaction and short-term impact. The program's monitoring and evaluation plan also includes an independent impact evaluation.[d]

Before full program rollout (in 2018), the program administrators used all the available data to adjust the program design to improve effectiveness. For example, moving from phase 1 to phase 2, the program administrators revised the communication campaign and training of community leaders on the community-based targeting approach to improve targeting and reach more families with children younger

continued next page

BOX 4.3 *(continued)*

than age five years, introduced a lump sum payment for the committee to cover the basic costs of targeting, and improved communication channels with the local agents in charge of supporting implementation activities. Preliminary results of the International Food Policy Research Institute's baseline survey (Hidrobo et al. 2015) highlighted the need for a case management system to address chronic undernutrition and anemia.[e]

a. The program's objective is the provision of a targeted cash transfer program to the chronic poor and vulnerable population and establishment of a basic system that could be used for other institutions to channel funds and services to the poor population. The Jigisemejiri program targets 62,000 households in a few districts (*cercles*) in all five regions of the south—Sikasso, Ségou, Mopti, Koulikoro, and Kayes—plus the district of Bamako. The program was implemented in three phases: phase 1 with a coverage of only 5,000 households, phase 2 to reach another 25,000 households, and phase 3 to extend coverage to 32,000 households.
b. Summarized in the UGTFS Rapport Annuel de Suivi des Activités, https://jigisemejiri.org/bibliotheque/.
c. See Statistics at https://jigisemejiri.org/statistiques/ and https://rsu.gouv.ml/portail/indicateurs/.
d. For the project document, see https://documents.worldbank.org/en/publication/documents-reports/documentdetail/384511468052440939/mali-emergency-social-safety-nets-project. For impact evaluation reports, see Hidrobo et al. (2015, 2018, 2020).
e. The case management system supports home visits by trained agents (social workers/nongovernmental organizations) to households with children younger than age three years to boost nutritional outcomes and promote access to health insurance (Régime d'Assurance Maladie, provided by Agence National d'Assurance Maladie).

monitoring to improve the program as it rolled out. According to Lindert et al. (2020), the main questions of interest are related to how well functions are being implemented and how well adjusted the tools in place for the program are to generate satisfactory results. Questions to be investigated include but are not limited to the following: Is program implementation running smoothly and as designed? Are there information and communication gaps that block people from registration? Is the intake and registration process effectively collecting all the information needed? How do vulnerable groups experience the process of intake and registration in the program? What obstacles do they face and why? Are they satisfied with the program implementation? Were staff properly trained? Are they satisfied with the resources made available? What percentage of intended population is registered? What percentage of the intended population is enrolled in the program? How long does it take from application to eligibility notification? How many applications are processed a

day according to quality standards? How many payments were processed during the last payment cycle? Assessments are to be conducted regularly using mixed methods and multiple data sources. The subjects may be the target population, applicants, beneficiaries, or the staff who work in the delivery system.

- Qualitative methods provide a descriptive approach based mostly on nonnumerical data to understand perceptions, observations, and social interactions. The main data sources are in-depth interviews, focus groups, and case studies.
- Quantitative methods rely on numerical data for monitoring and identifying trends. The main data sources are program-specific surveys of potential or actual clients and program administrative data, often triangulated with wider population data from household surveys. The program-specific inquiries may elicit both clients' and administrators' perceptions of certain aspects of the program that affect its implementation and outcomes, such as compliance with quality standards, client understanding and satisfaction, and access for vulnerable groups to services, among others.
- Program administrative data can be a rich resource as part of regular monitoring and evaluation for measuring program outcomes, including targeting. For example, dashboards can be created to generate information on applicants, time for processing applications, grievance and redress performance, and information management and control mechanisms. Similarly, administrative data can help in understanding program staff caseloads and program staff work quality, with indicators such as staff turnover rates, budget execution, and compliance with operational procedures and program rules.

Planning and Adapting Delivery Systems for Crisis Response

The social protection sector is increasingly called on to ensure that its delivery systems are well prepared to handle disaster responses. Because there is a premium on the speed of response in crisis, there is a premium on preparedness. Bowen et al. (2020), OPM (2017), and UNICEF (2019) highlight factors that enable social protection systems to be responsive to shocks and deliver effective shock response. Many parts of delivery systems for programs focused on equity and opportunity for normal times can be the foundation of delivery systems for crisis response, but there are some specificities to crisis response and some extra considerations. This section examines those.

Covariate shocks can be classified into two broad groups according to their degree of predictability:

- *Greater predictability.* This category includes most of weather- and climate-related disasters (for example, droughts, storms, wildfires, and floods) that occur periodically and may be increased in frequency or severity by climate change and degradation (depletion) of natural resources. They are often referred to as "hydro-meteorological" shocks, and some can lead to mass casualties and/or major damages to property and disruption of means of livelihoods, roads, and the normal way of life of the people in the affected areas.
- *Low predictability, infrequent, or unpredicted.* This category includes all geo-physical disasters like earthquakes, tsunamis, volcanic eruptions, and dry mass-land movements, which are not caused by climate change; economic crises, such as the 1997 Asian Financial Crisis and the 2008 Global Food, Fuel and Financial crisis; epidemic crises (for example, Ebola, H1N1/swine flu, H5N1/avian flu, and COVID-19); and insect infestations (for example, the locust swarms of 2020 centered in the Horn of Africa but extending far beyond).

Shocks can be serially correlated. For example, floods can be followed by an increase in illness, malnutrition, and deaths due to the transmission of water-borne diseases.

Preparing to Handle Eligibility Determination for Social Protection Responses to Natural Disasters

A first step in planning for agile responses to natural disasters is assessing hazards ex ante. Hazard assessments should consider the risks related to the most likely or most severe hazards that might affect people or their assets, the expected impacts on consumption or income, and the distribution of impacts across regions or the welfare distribution (see, for example, World Bank [2021] for a stress-testing toolkit and Hill and Porter [2017] and Porter and White [2017] for applications). Past experiences are an important source of information to inform adaptation and changes that could make the social protection system more responsive and adaptive to shocks. For example, Mori et al. (2020) show that the following occurred during 2010–15. Ethiopia experienced seven droughts, affecting a total of 43 million Ethiopians (almost 40 percent of the population), placing them in food insecurity and deepening poverty levels. In the Philippines, during the same period, 131 storms hit the country, killing 19,000 and making 321,000 homeless; Typhoon Haiyan alone increased the number of people in poverty by a million people (Bowen 2016). Pakistan was hit by 14 major earthquakes, resulting in 74,000 deaths, 132,000 injuries, and more than

5 million people made homeless. Based on such past experiences, countries can develop plans and frameworks to reduce the impacts of shocks through preventive measures such as on improving program delivery chains, designs, data systems, knowledge management activities, intersectoral coordination, and so forth, and consequently be ready for a faster and better response (Bowen et al. 2020; UNICEF 2019).

With some notion of disaster profiling in hand, countries can plan what sort of programming they might rely on for recurrent emergencies and how delivery systems would support that. Responses may be multi-layered calling on different programs and options depending on the severity of the shocks.

- *Continuously operating programs with built-in flex can carry part of the load even with no specific disaster-related triggers.* Entitlement-based social assistance programs are built to expand when needs are higher. To make these programs work, they need open registration systems, so anyone suffering a loss in income can qualify, and budget provision that guarantees that all who qualify get access. Unemployment insurance programs similarly work automatically, expanding benefits in downturns. Although they are not usually much discussed among disaster responses, such programs can be important. For example, Deryugina (2016) looks at US counties that were hit by hurricanes. He estimates that about 80 percent of the fiscal support in the 10 years following flows through the US regular safety net (unemployment insurance, means-tested social assistance, and public health programs), and the minority of funds flows through the disaster-specific provisions of the Federal Emergency Management Agency. During economic crises, such programs are an even more natural fit. In the 2009 financial crisis, in Eastern Europe and Central Asia, the first wave of response came through increased unemployment claims and the second through the region's last resort social assistance programs. However, many of these programs had eroded in eligibility threshold prior to the crisis, which limited some their responsiveness, and some countries took action during the crisis to reform their programs (Isik-Dikmelik 2012).
- *Continuously operating programs can have preestablished disaster-related triggers.* Ethiopia's rural Productive Safety Net Program and Kenya's Hunger Safety Net are iconic examples of having well-established systems that trigger expanded caseloads based on climatic data. In Ethiopia, the number of people covered under the rural Productive Safety Net Program and the number of months of benefits per year has been adjusted regularly in response to drought. The trigger has been the twice-annual Humanitarian Requirements Document. To improve the automaticity, speed, and geographic differentiation of response, the government is planning to switch to triggers based on an early warning system and

incorporate elements such as agrometeorology, crop data, livestock data, market prices, poverty, and population (World Bank 2020e). In Kenya, the Hunger Safety Net program already has a census of households in the covered districts and the information to issue payments directly to their accounts when the predefined triggers are hit. These are based on the vegetative condition index, and depending on the level of the index, expand the number of households, level of benefit, and/or frequency of payment (UNICEF 2019).

- *Even when not planned so systematically, emergency top-up benefits for social assistance beneficiaries in affected areas are a common policy response*, and sometimes eligibility is expanded. For example, in response to Typhoon Haiyan, top-ups to the Philippines' cash transfer program (Pantawid Pamilyang Pilipinoa Program, or 4Ps) reached beneficiaries within a month of the storm (Bowen 2016; Pelly, de Wild, and Inarra 2015).[16] In response to the 2020 locust plague in the Horn of Africa, Djibouti worked through the existing Family Solidarity Program with a scale-up of about 15 percent over the base caseload, and the expansion focused on the areas that were most affected (World Bank 2020d).

A relatively new frontier is collecting data at the household level ex ante to predict vulnerability to natural disasters. Geotagging households means that many spatial data on risks of shock exposures can be cross-referenced. (Geotagging is becoming more standard for allied purposes of mapping service provision and identifying households unambiguously.) In the Dominican Republic, the information collected for its social registry (SIUBEN) also includes vulnerability to climate shocks within its database, with information on three dimensions: (1) housing characteristics (walls and ceiling), (2) estimated income, and (3) proximity to a hazardous natural element (river, stream, or ravine). Pakistan is including data on climatic vulnerability in its new PMT, while also making efforts to provide geographic coordinates for all registered households (UNICEF 2019). Colombia's System for the Selection of Beneficiaries for Social Programs (SISBEN IV) moved in this direction as well. For the first time, it undertook the geolocalization of all households and added to the household questionnaire a module assessing household exposure to natural disasters (World Bank 2020c). Chapter 6 discusses how to incorporate risks into a PMT in more detail.

When assets are destroyed or welfare rankings much changed, there may need to be a postdisaster eligibility assessment geared to the areas affected. These ex-post assessments inherently add a step and expense to response times, but if prior systems are set up well, the assessments can be done relatively expeditiously. From a social protection angle, knowing the characteristics and profile of the population and its vulnerabilities

ex ante by expansion of registries and interoperability of data systems can help in preparing the ground for emergency response. In Chile, a country with high vulnerability to natural disasters, the integration of social protection and disaster risk management was prompted by citizens' complaints that in preceding disasters there were too many mistakes in the process of gathering data, extensively long compiling and processing time, and the aid was not designated accordingly. Thus, in 2015, the overall disaster risk management system was revised. The role of the Ministry of Social Development is to coordinate the application of ex-post, household-specific needs assessments. Accordingly, it developed the Emergency Basic Fact Sheet (Ficha Básica de Emergencia, or FIBE) with attendant mobile apps and tools. FIBE is linked to the Chilean National Social Registry, reducing the data collection time substantially according to the ministry, as most of the basic information is prepopulated in the FIBE information system. As a metric of efficiency, data collection in the response to the Coquimbo earthquake in 2015 took 27 days using the new FIBE, in contrast to the response to the Tarapacá earthquake in 2014, which took 115 days (Beazley, Solorzano, and Barca 2019).

Thinking about Economic Shocks

Widespread economic shocks present their own challenges to social protection systems. Shocks such as the Asian Financial Crisis in 1997/98, the global financial crisis in 2007–09, food price surges in 2007/08 and again in 2011, or the economic consequences of COVID-19 manifest in worsening income of those already poor, loss of earnings and/or jobs in the informal sectors, and some formal sector job loss. In general, physical assets are not destroyed by economic shocks, although savings may be wiped out by inflation, devaluation, or declines in financial markets. Moreover, households and businesses may have to divest themselves of assets, especially more liquid assets, to survive in the short term, thus lowering their future earnings. The effects are usually national in scope rather than geographically delimited, and often the duration is initially uncertain. This means that the one-off, geographically focused benefit often used for natural disaster response will be less appropriate for economic shocks, and it puts a higher demand on delivery systems that can adjust over time, possibly those that can follow changes in welfare dynamically.

The COVID-19-precipitated crisis had a more sudden and precipitous onset than usual for economic shocks, one that called for lightning speed response and thus put an even heavier burden on delivery systems than usual. The desire to move extraordinarily fast and with little social contact implied favoring the use of existing data or virtual application processes

rather than traditional face-to-face interviews or field work, which was accomplished in different ways:

- Sixty-eight countries made use of top-up benefits for those in existing programs.
- To achieve the increases in coverage desired, some countries mounted enormous new digital registration efforts. Several middle-income countries (including Brazil, Jordan, Morocco, Namibia, Pakistan, South Africa, Thailand, and Turkey) used their social registries or other existing government databases to expand coverage hugely from base levels, essentially inviting many to apply and ruling out people with formal sector incomes or recorded assets above certain levels, or those in receipt of various government benefits. Brazil initiated a new Emergency Aid program with online and mobile application technology, which eventually reached about 68 million people, one-third of the population (Gentilini et al. 2020, v14). The government information and communications technology firm DataPrev cross-checked claims against Cadastro Único and some 20 constantly updated databases, including tax, social security, public employment, and Brazilians resident abroad (World Bank and FCDO 2021). Pakistan's Ehsaas Emergency Cash program reached about 45 percent of the country's population, relying on top-up benefits to its ongoing PMT targeted program and giving benefits to those in the social registry who were above the usual threshold PMT and, because the registry was old, via new applications screened for a list of exclusions.
- Some countries managed to innovate using untraditional databases. Togo happened to have a very recent biometric voter registration database from elections in February 2020, which (unusually) contained information on occupation as well as location. The government managed a fully digital registration and payment process to issue payments to residents in areas affected by lockdowns and with occupations in the informal sector because they were presumed to have or be at risk of significant income loss (Boko et al. 2020). In Guatemala, the government introduced an emergency cash transfer, Bono Familia, during three months (1,000 quetzals, or US$130, per month/beneficiary). The program gave benefits to 2.6 million households (80 percent of the population) consuming less than 200 kilowatt hours for areas with electricity and made provision for 0.2 million more lacking connections.
- Some countries used more traditional methods. For example, the Philippines both topped up benefits to those in the 4Ps program and undertook substantial new registrations with face-to-face processes and manual payments, which was a rather slower process. Almost all 4Ps beneficiaries were able to receive the Social Amelioration Program first tranche top-up benefits through the already existing digital channel (for

example, cash cards) by April 5, 2020. In contrast, about one-quarter of the target group to be served by the new registration had received their first tranche of benefits by April 25 (Cho, Avalos, et al. 2021). The difficulties in manual processes led the government to digitize the processes for a second wave of responses (Cho, Kawasoe, et al. 2021).

Countries with higher base rates of foundational IDs, financial inclusion, mobile penetration, and interoperable government databases or social registries were better able to provide response more quickly. Although most countries made policy announcements quickly after the World Health Organization officially declared COVID-19 to be a pandemic in March 2020, getting cash into the hands of the population was sometimes slower. Palacios (2020) calculates that among 66 programs that announced responses involving new beneficiary intakes (rather than those with pay-outs of top-up benefits), about half had managed to make payments by the end of June 2020 (see figure 4.2). The ability to roll out quickly was much greater where the building blocks of foundational IDs and bank accounts were widespread. High-coverage social registries were partly helpful, but

Figure 4.2   Mapping the Roll-Out of Coverage of Identification (IDs) and Financial Inclusion
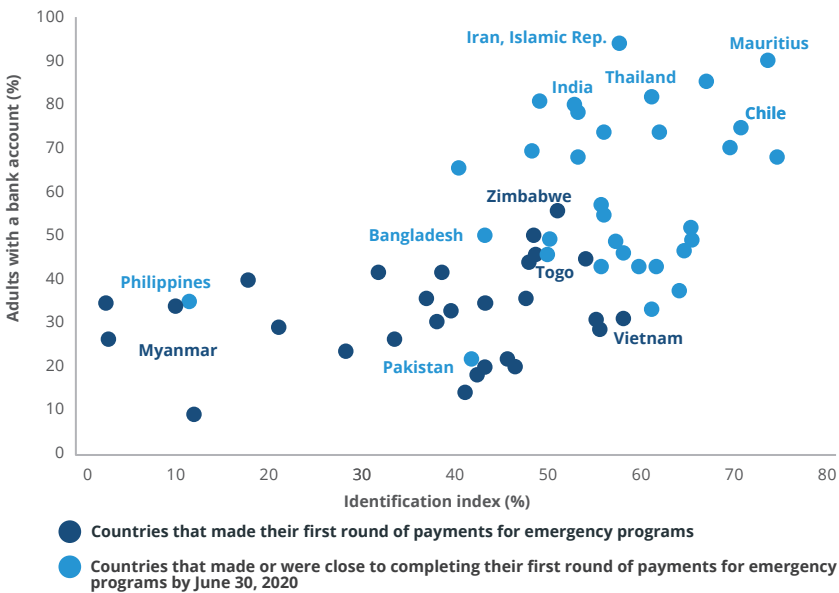
Source: Palacios 2020.

since some had data that were several years out of date and none had post-crisis data, many countries supplemented these with new application processes, which were built on the existent infrastructure of large program data systems and social registries in most cases.

The emergency responses to the COVID-19-induced economic crisis echo a refrain learned in past crises—when there is urgency of action, what might not have been imagined possible can be done, albeit rarely perfectly. Many countries rolled out programs in record time to record numbers of clients, a fact more impressive because it was accomplished while many aspects of logistics were impaired by closures of public offices, public transport, and other supporting services. Enormous good was done, although in most countries it was not enough to match the scale of losses. But the speed meant that systems could not be fully tested and every wrinkle ironed out before going to scale. It made it hard to overcome all the usual barriers to perfect inclusion. Thus, to varying degrees in different places, challenges were seen at various scales. In some cases, these were reasonably resolved with quick troubleshooting, in other cases, they were less so. Online portals could not always keep up with the onslaught of applications, although usually they caught up eventually. The financial service providers could not always keep up with the pace of payouts and ran out of cash or saw long lines. The issues of digital divide, lack of financial and basic literacy, not speaking the official languages, people whose fingerprints do not scan well, and all the usual challenges were encountered, with responses sometimes impaired by the speed and/or low-contact way of working. Often emergency programs are done so fast and pass so quickly that neither impact evaluation nor much real-time monitoring can be done. Some reports of accounts of the kinds of issues and real-time troubleshooting that occurred are emerging, and many underscore the need in the long run to address some of the fundamental constraints in delivery systems that limited responses. For example, Gelb and Mukherjee (2020) and Palacios (2020) look at global responses; SPACE (2021) provides a bibliography of regional and country-specific materials; and others study individual countries, including UNECA (2021) for Namibia; Nishtar (2020) for Pakistan; and Cho, Kawasoe, et al. (2021) for the Philippines.

Client Interface: The Interaction between People and Institutions

There are many modalities or touch points for client interface with institutions. Various interactions occur in person with frontline workers. The location of the interactions can be people's homes (via home visits by mobile teams), temporary community sites, permanent local offices, or

specific points of service (including payment providers). Instead, interactions may occur digitally, via call centers, self-service kiosks in public spaces, mobile devices, personal computers, and so forth. Ideally, there are multiple channels to serve clients with different needs and constraints, with easy accessibility being the watchword—convenient hours; staff who are conversant in pertinent languages, well trained, able to handle the processes requested, and supported with adequate information systems; physical access where pertinent (including for the mobility impaired); and with reasonable amenities (protection from rain and sun, access to water and sanitation, and safe places for accompanying children).

In the era of COVID-19, concern about touch points went from the metaphorical to the physical/epidemiological, sparking a wave of adjustments to delivery systems. The use of digital delivery garnered huge impetus, with many countries pushing advances on several fronts. Some rolled out or reinforced electronic application processes. Many pushed harder on digital payments—changing regulations, opening accounts for new beneficiaries, waiving fees for accounts or digital transactions, raising limits on transactions, and so forth (Gentilini et al. 2020, v9). Where physical contact was still needed, countries generally declared social services centers as essential services and left them open, simultaneously making a range of efforts to reduce the chances of spreading the virus, for example, through increased sanitation measures, and spreading payments across more days of the month or more payment providers to reduce queueing and facilitate social distancing.

The impetus that COVID-19 responses gave to digital solutions reinforced the secular trend and may increase convenience for some social assistance clients, but there is a long way to go to bridge the digital divide. The countries that were best able to use digital enrollment and payment procedures in COVID-19 responses were those that had preexisting conditions, such as high-coverage foundational IDs, high-coverage social registries, linkable information from social security registries, and existing account-based or digital payments or amenable legislation that enabled a switch. Even so, while evidence on the extent of errors of exclusion from COVID-19-related digital systems is still scant, there is a great deal of evidence that documents the digital divide pre–COVID-19 (for example, ITU 2020; World Bank Group 2016). For example, in low- and middle-income countries, women are 8 percent less likely than men to own a mobile phone, 20 percent less likely to own a smartphone, and 20 percent less likely to use the internet on a mobile (GSMA 2020). Gender gaps in phone ownership are highest in South Asia—over 30 percentage points in India and 20 percentage points in Bangladesh and Pakistan (Bashir et al. 2021). Thus, as the motivation for digital services reverts from contagion control to efficiency and accountability, countries and programs will need to

continue to help to raise coverage of IDs, payment accounts, and digital services for social assistance clients.

Maintaining some human help for potential registrants who lack access or acumen to use the digital platforms can be an important strategy. Physical offices, public access kiosks, and mobile outreach services will be needed for the many who do not have effective access to digital services—because internet services lack coverage or quality or are unaffordable, the potential clients do not own devices, they lack experience and agency to solicit government services digitally, or they face more basic barriers of literacy and language. The gaps in access in low-income countries are well known. Worldwide in 2019, 51 percent of people used the internet (ITU 2020). In African countries, the poorer 40 percent of the population is only one-third as likely to have access to the internet as the upper 60 percent; men and youth use the internet much more than women or older generations, and so forth. Similar gaps are found even in Europe—citizens in the top 20 percent of the income distribution in the most connected European Union (EU) country are 45 times more likely to use e-services than those in the bottom 20 percent in the least connected EU country (World Bank Group 2016). In South Asia, usage remains low, at around 10 percent in Bangladesh and 40 percent in Sri Lanka (Bashir et al. 2021).

On the institutional side, the interface with clients may be the local staff of the central social protection ministry (or wherever in the institutional landscape the particular program may be); municipal staff who carry out some program functions in joint implementation agreements with the central agency; or outsourced providers working on contract, such as not-for-profit social service providers or firms doing survey enumeration. In Mexico, the PROGRESA-Oportunidades-Prospera program was run with federal staff working throughout the nation. Using local governments to implement federal programs is more common. In China, the *dibao* program is administered by local civil affairs bureaus, with responsibility for determining eligibility, thresholds, beneficiary selection, and transfer payment amounts. In Brazil, the municipalities carry the workload of getting households entered into the social registry that serves as the gateway to dozens of programs. In Ethiopia, the district councils are important actors in the rural Productive Safety Net Program. Even in advanced countries such as the United States, local levels carry out important functions. Each arrondissement in France and county in the United States has at least one center with dedicated social workers ready to receive applications, conduct home visits, and give advice to people in need, as well as help with administrative procedures and refer clients to services that can meet their needs. Hence, the physical location of such services is important and spatial analysis can be useful in determining where to place capacity (see box 4.4).

BOX 4.4

Geographic Information Can Help Diagnose Bottlenecks in Program Access That Lead to Errors of Exclusion

For Croatia, Azevedo (2017) compares poverty maps with spatial patterns of the availability of social welfare centers (SWCs) and spending on core social protection programs at the municipal and city levels. In Croatia, any individual who may require a social benefit must first apply at a social welfare center. Application for a program only requires a one-time visit by an applicant, and all benefits may be sent via post, bank deposit, or picked up in person. In a such a small country and with a single trip required, this is a context in which the location of offices would have the least power to influence participation, and yet, even in this context it does, and the government has been attentive to minimizing the issue.

There are 80 SWCs in Croatia, meaning that one center covers several municipalities/towns, close to seven on average. The farthest distance from the center of a municipality to its closest SWC is 25 kilometers. This implies that when someone wishes to apply for a benefit, he/she must travel less than 25 kilometers, at least once. The average citizen must travel 1.8 kilometers to the closest SWC branch. The shorter is the distance that an individual must travel to request assistance, the more likely it is that the individual will seek assistance. Ideally the centers should be closest to the poor to minimize the burden of seeking assistance. The analysis shows that social welfare centers are closer to the poor, but they are not necessarily closer to places with higher poverty rates. Individuals in and around the city of Zagreb must travel the shortest distance to get to a center, while individuals in some of the poorest places in Croatia must travel the farthest. Most of the poor are in places that are more densely populated, and, consequently, the centers are closer to where the poor are.

The study finds that spending is reasonably well targeted toward the poorest municipalities. There are relatively more beneficiaries of social programs living in municipalities and towns with higher poverty incidence and poverty depth than those living in richer areas. Further, relatively large amounts of social benefits are distributed toward the poorest local units. However, the distributional outcomes differ significantly across social programs, with the guaranteed minimum benefit found to have the most progressive incidence.

Each of these arrangements requires clarity of rules, authority, budget flows, and, ideally, compatible incentives between the parties. These inter-agency arrangements are largely beyond the purview of this book, but they are important nonetheless. For example, if local authorities can select beneficiaries and bear none of the funding burden, they will have incentives to exercise any latitude they have in favor of increasing local receipts, possibly to the extent of introducing errors of inclusion. Conversely, if local authorities are expected to do work for a federal program but not given adequate resources for staffing, IT, transportation, and outreach, there are likely to be errors of exclusion.

Where formal administrative capacity has been lacking (not yet built) in pertinent government agencies, communities are sometimes called upon to fill the void. Communities sometimes serve as deciders in household-specific determination of eligibility (a topic covered in chapters 5 and 6), but they can also serve in helpful ways even when full community-based targeting is not among the targeting methods used. Communities can play several different roles. Sometimes as delivery systems are built and the focus shifts from single programs or benefits to multifaceted programs and social protection systems, community participation is maintained but channeled into specific areas of comparative advantage and/or community partners are transformed into local administrative agents.

Community members such as community spokespersons or mother-leaders serve as important elements in communication and supplement the government's capacity in various programs. Mexico's PROGRESA program started the model, which was picked up in other countries. In Colombia, mother-leaders are selected by their communities and receive some training by the programs. They help the community to understand the program and its rules, when and how to receive payment, and the procedures to file a grievance, and sometimes they help to amplify the messages received in behavioral change sessions. Pakistan's Benazir Income Support Program also uses its extensive network of mothers' groups as outreach agents. Similarly, a community-based outreach model in the Republic of Yemen's Social Development Fund focuses on using existing networks in the community for outreach.

Building the network for client interface is an ongoing task, more advanced in some places than others, probably still incomplete in most countries, and in constant need of attention. Some countries have made significant efforts to build a system where they had little a few years ago (see examples at the end of chapter 5). Especially in countries that are newly building programs and sometimes in countries with more established ones, it is typical for the networks to fall short of ideal. There is much to be done to improve social protection delivery systems, which will help with a variety of aspects of program success, including ensuring that the program reaches those it means to serve.

Data Systems and Their Role in Supporting Eligibility Determination and Recertification

Data are at the heart of decisions about eligibility and important for managing all the steps in the delivery chain; thus, good data systems can improve the targeting and impacts of social protection programs. Data systems require the support of those with specialized skills, but their basic functions are intuitive: to collect or assemble correct data and store them in a way that makes their use both easy and secure. This section is a brief reminder of the issues, with references to the wider and more technical literature on the topic.

Data Collection or Aggregation

Data may be collected new from applicants or gathered from existing data that are already accessible by the government. For the assessment of needs and conditions or eligibility determination, new data collection comes in the form of the traditional face-to-face interview—in a program office when a person comes to apply or in a survey sweep when a program or social registry goes to the field to collect data in people's homes. The modern variant is the virtual interview by phone, app, or online form.

Although eligibility determination is particularly heavy on data and inference (an issue covered in chapter 6), data are collected and used throughout the whole program cycle. Good outreach, GRMs, and error and fraud detection subsystems all require the collection and use of large amounts of data. Data may be sent to agencies where recipients of income support programs are referred to other programs for which they may be qualified. Of course, data with respect to the payments to be made will need to be conveyed to payment service providers. Data may be drawn from other agencies not just for eligibility determination, but also to verify that claimants have fulfilled any co-responsibilities, for example, registration for public employment services or attendance at training, school, or health care.

In many instances, data collected from applicants are the only or main source of program data. This is particularly true in low-income countries with large informal sectors and few administrative databases that cover only a minority of the population, with an emerging but incomplete legal and regulatory framework, lacking a secure and trusted architecture for exchanging data among different ministries and agencies, and with an insufficiently resourced public sector (for example, lacking the capacity to develop or contract IT expertise), or with insufficient human resources that can analyze and make use of the data (Lindert et al. 2020). These difficulties notwithstanding, programs can develop efficient data systems by collecting and digitizing the information from all business functions

(from outreach to payments), developing data quality control routines, and using the different data subsystems from each business function in an integrated way to detect data gaps, incompatibilities, outliers, or changes that trigger changes in other parts of the data system. To improve the completeness, timeliness, and accuracy of data, programs can open multiple channels of communications with beneficiaries, to ease the cost of updating and correcting their data, and integrate with relevant information from other stakeholders (for example, payment providers or health, education, or employment agencies).

The advantages of collecting new data for eligibility determination are two. First, the program itself has agency as it does not need to wait for the surrounding data ecosystem, which leads to the second advantage—the program can work with people or aspects of their welfare that are not recorded in data systems that are accessible to program administrators.

When data are being collected anew in interviews, data quality can be supported with lessons from household surveying and process evaluations. Because there is a sound literature and practice for surveys,[17] this section does not dwell on the topic but provides a few examples as reminders of its importance.

- The use of computer-assisted personal interviewing, which is commonly conducted on tablets or smartphones, can facilitate more efficient quality control through automated checks (examples include range limits, logic checks, coverage errors triggering supervisor visits, and location flags to prevent duplicates and other errors). This technology also reduces cumbersome processes, like double data entry. For a small example of the power of these kinds of tools—in performance diagnostics for Colombia's social registry (the System for the Selection of Beneficiaries for Social Programs, or SISBEN III)—DNP (2016) notes substantial errors in the handwritten addresses and ID numbers, problems for which it posed solutions for the implementation of SISBEN IV by instituting a range of checks in the computer-assisted personal interviewing technology.
- Video is also becoming a more common tool that can help standardize training, allow new people to be trained as they enter the social assistance workforce in a continuous rather than cohort approach,[18] and allow staff to brush up on their skills as needed. A study in the Philippines (Velarde 2018) shows the problems of the former "cascade training" or "training of trainers" approaches commonly used prior to the advent of affordable video. The study documents that loss of key concepts occurred at each stage of the cascade, resulting in additional time spent on supervision and correction of errors, problems that were reduced by the videos.

- Digital training and quality control have been successfully tried in India in the Aadhaar (unique digital ID) program.[19] To enroll the 1.24 billion people, Aadhaar has used a mix of private and governmental actors, organized hierarchically into registrars, who selected enrollment agents, who in turn organized enrollment centers. The whole process was decentralized to 628 enrollment agencies with more than half a million enrollment operators. To ensure that all enrollment operators are properly trained and produce valid enrollments, Aadhaar has organized a multimode system of training, testing, and certification standards. Given the size of the country and program, these functions were delivered not only through traditional instruction manuals and face-to-face training, but also digitally: training in e-documents and videos and testing and certification were computer assisted.
- Real-time supervision of data collection is important. For example, Pakistan's National Socioeconomic Registry update process closely monitors coverage through multiple sources, including directly from its regional control center in real time, using geotagging of questionnaires as well as indirectly through an independent operational review firm. Based on this information, it regularly contacts enumerator teams to resolve exclusion issues, particularly in harder to reach and more vulnerable areas. It also conducts early monitoring and household surveys to identify constraints to the poor in accessing desk-based centers (for example, placement of centers, costs, and waiting times) and error, fraud, and corruption in the process.

New data collection is costly and time consuming, imposing administrative costs, transaction costs, and sometimes stigma on applicants, which may result in biased information.[20] Thus, there is a strong impetus toward using, to the extent possible, data that have already been collected. Instead of asking the applicant about their income or assets, the social assistance program asks for the client's consent to use other records and then draws on data held by other agencies—for example, those that track social security contributions, registrations for automobiles, and so forth.

The buzz in data collection is about increasing the use of data that are already available. The society-wide secular trends of the falling costs of computers and communication and the exponentially increasing use of mobile technology in communications and commerce are key drivers. Moreover, as countries move from a model of individual, island-type social protection programs toward an archipelago of programs, often through the development of a social registry or an integrated system, the emphasis shifts further toward the use of outside-the-program sources and interoperability between different programs' data systems. (Box 4.5 provides examples of these.)

BOX 4.5

Development of Interoperable Data for Eligibility Assessment in Social Assistance Programs: Examples from Chile and Moldova

The two examples in this box make two complementary points. First, the case of Chile presents a continuous and gradual improvement in the data systems of social assistance programs and the development of interoperability at all four levels discussed in the text—legal, organizational, semantics, and technical. Second, the case of Moldova presents a leapfrogging reform that changed the data paradigm from paper to digital within a short time and at a relatively low cost.

The evolution of Chile's information system supporting eligibility determination illustrates the shift in good practices with the digitalization of public services. The first application form, Ficha CAS in the 1980s, was a paper form administered by social workers who collected self-reported data from applicant households. A second version of the Ficha, which was used until 2006, continued to collect self-reported data that were then digitized. The move toward digital data capture and collection gained speed during the 2000s and 2010s, and the number of data sets that fed into the Integrated Social Information System increased as well. The latest version of the Integrated Social Information System, the Registro de Social Hogares (RHS), integrates data from 43 public sector agencies with some self-reported information on informal income, occupation, housing, education, health, and family composition that applicants to social protection programs can supply online or through local municipal offices. The RHS helps determine eligibility for 80 public programs. The RHS is dynamic: most of the administrative data are updated monthly. The development of the RHS and its predecessor was facilitated by the development of a strong data protection framework. The right to privacy of all people is recognized, protected, and guaranteed by the Chilean Constitution (Article 19); a 2018 amendment established the protection of personal data as a constitutional right.[a] A personal data protection law adopted in 1999—well ahead of the European Union (EU) General Data Protection Regulation, for example—established the purpose limitation principle, the requirement to protect sensitive personal data, the rights of the data subjects[b] as well as the right to receive damages. The legal framework is complemented by sector-specific laws and implementation

BOX 4.5 *(continued)*

regulations, as well as a law on sensitive data (2011, amended in 2012). The enforcement system relies on civil courts. A draft data protection bill includes the creation of a dedicated data protection agency. On a technical level, the development of RHS was enabled by a whole-of-government approach to move toward digital, which is spelled out in the Digital Transformation Strategy (2019, https://digital.gob.cl /biblioteca/estrategias/estrategia-de-transformacion-digital-del -estado), with its three objectives: to improve public services for citizens and businesses, to engage in evidence-based policy making, and to mainstream the digital transformation across government and the economy (Silva et al. 2018; World Bank 2020f).

Moldova put together an interoperable data system fast and at relatively low cost. The Moldovan Social Assistance Automated Information System (SAAIS) was designed to increase the efficiency of social

Figure B4.5.1   Moldova: SAAIS Business Processes
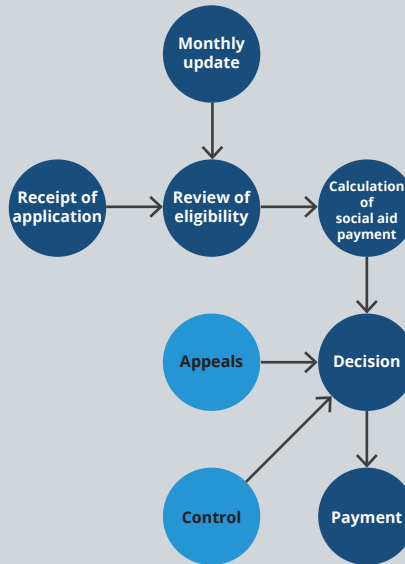
Source: Sluchynskyy 2019.

BOX 4.5 *(continued)*

workers, district managers, and ministry staff (figure B4.5.1). It is used by 1,300 users in more than 950 offices. Development of the SAAIS began in 2011, with a first operational version completed within 18 months at total software costs of $1.3 million. The development benefited from an adequate legal and institutional enabling framework. The development of web services for data exchanges was quick (weeks), costing less than $10,000 each, but the process of finalizing the service agreements between agencies took months. A key feature of the system is automated verification of information provided by the applicants for social assistance against data contained in other administrative databases. When a social worker inputs a citizen's application, the following data are automatically pulled from other agencies using the system of web services: Population Registry (data on the applicant and members of his/her family), National Transport Registry (vehicles registered in the names of the applicant or family members), Employment Agency (employment status of the applicant and family members and information on unemployment benefits and registered rejections of proposed jobs), National Office of Social Insurance (pensions and benefits provided to the applicant and family members), National Cadastre Registry (land plots and immovable property registered in the name of the applicants and their family members), and Border Guard Service (to establish whether the beneficiary is currently in the country). The system also provides automatic monthly reconciliation with all agencies before the payment lists are prepared. As a result of reconciliation, all statuses are automatically adjusted based on actual data received (Sluchynskyy 2019).

a. The Constitution ensures the following for every person: respect and protection of private life and the honor of the person and their family and the protection of personal data. The treatment and protection of these data will be put into effect in the form and conditions determined by law.
b. These include the following: the right of modification, if the personal data are erroneous, inexact, equivocal, or incomplete; the right to block processing when the individual has voluntarily provided his or her personal data but no longer wants it to be processed; the right of cancellation or elimination of expired data; the right to access their data for free; and the right to oppose the use of their data for advertising, market research, or opinion polls.

A program's data system's ability to communicate by exchanging data, so the information is understood by the receiving agency and subsequently used for its own business purposes, is called interoperability. Interoperability encompasses different levels of integration: legal (data exchange is allowed), organizational (business processes and governance are aligned to facilitate data exchanges), semantics (data definitions/metadata are compatible),

and technical (security protocols, transmission protocols, and various standards). In practice, data systems can be interoperable on all these levels or only selected levels.

Interoperability does not develop in a vacuum. Various enablers and safeguards facilitate interoperability and trusted data sharing (World Bank 2020f). One level of these encompasses regulations and institutions that go well beyond those over which any particular social protection program has agency: a policy and regulatory environment that defines and enacts rights over data, robust and resourced institutions capable of enforcing the rules while also offering citizens responsive and effective redress, technical architecture to standardize data sharing within government while giving people more controls and providing transparency of data flows, capabilities inside and alongside government to analyze and make use of data, and an active civil society and informed populace that can effectively use data and keep governments and companies accountable. Another level of enablers comes from more technical investments that enable data sharing and data security: (1) interoperable databases that are accessible to and used across government agencies for sharing data; (2) e-services portals that allow citizens to access government services and individual data portals that allow people to aggregate, store, and share data; and (3) inclusive digital platforms such as digital identification that ensure that all people are participants in the digital economy. All these factors are part of the data ecosystem and influence how extensively an individual social protection program can use external information.

The use of data external to the program comes with benefits but also risks. Among the benefits, interoperability can reduce transaction costs to the applicant, saving the time and hassle of supplying the same information time and again to different government agencies. Administrators can find efficiency gains in data quality and accuracy, reducing duplications and errors and improving transparency, while lowering administrative costs as the developed data system reduces the cost of repetitive data collection. Investment in the data system helps to improve social programs, including targeting, financing, and planning, by providing better coordination in identification of target groups and coordinating social programs. Among the risks, there is potential perpetuation of some inequalities and bias (exclusion) against certain groups and issues related to data privacy and security, which have implications for human rights, if not well attended.

Data Privacy and Data Protection

The transition toward e-Government is influencing the way information is collected, managed, and reported by governments, including

social programs. In the not so distant past, information collected by social programs was mostly in paper form and not very extensive. As such, it did not raise substantial issues of data privacy[21] nor trigger complex responses to ensure data protection.[22] With the generalization of digital information flows and increased use of digital technologies in social programs—ID systems, biometric data, interlinked data for eligibility determination, and digital payment systems—concerns about data privacy and data protection have become important (box 4.6).

Data privacy risks can arise from any activity that collects, stores, or processes personal data. Among the information collected by social programs, most concerns are related to management of personal data (any information relating to an identified or identifiable individual), other personally identifiable information (information that permits the identity of an individual to be directly or indirectly inferred, or any information

---

BOX 4.6

Why Should Social Protection Practitioners Care about Data Protection?

Social protection systems process (collect, use, store, and disclose) the personal data of applicants and program beneficiaries. These data need protection. Why?

- If personal data are not adequately protected, the data subjects' right to privacy may be violated and individuals may suffer material, physical, or symbolic harm.
- Data protection is essential to create trust among social protection authorities, their staff, and clients. Lack of trust may restrain the access of vulnerable populations to social protection services and benefits, as they may fear that sharing their personal information will lead to harm, discrimination, stigmatization, or surveillance, among other risks.
- For social protection practitioners, compliance with organizational or legal data protection and privacy frameworks is important to avoid penalties.
- Social protection and privacy are human rights and, therefore, interdependent. This means that one needs the other to be fulfilled. Both are equally important.

Source: SPIAC (forthcoming); Enabling Digital & GIZ (2020).

---

that is linked or linkable, or may be attributed, to that individual, including street address, email, telephone, IP address, geolocation, biometric, or behavioral data), and sensitive data (sexual orientation, membership in an ethnic minority group, trade union, and so forth). As social programs deal with socioeconomic data as well as data sourced from multiple government systems through interoperability and data integration protocols, the social protection sector must ensure that a process for protection of personal data is in place.

Data protection legislation and institutions should protect the use of personal information against risks such as exposure of personal data, data and identity theft, discrimination or persecution, exclusion, unjust treatment, and surveillance. Protecting personal data is a critical aspect of the design of systems enabling a relationship of trust (Alston 2019; Bashir et al. 2021; Ohlenburg 2020; Sepulveda Carmona 2018. Data protection legislation typically offers protections for applicants and beneficiaries, such as the rights to object, access, rectify incorrect records, erase, and restrict processing to what is minimally required to operate the program, as well as the right to notification in case of processing, breach, and so forth.

The United Nations Personal Data Protection and Privacy Principles[23] are important for the delivery system of social protection programs. The United Nations High-Level Committee on Management 2018 highlights the importance of having data processed in a fair and legitimate manner, considering the person's consent and best interests and processed and retained consistently with specified purposes. In addition, it highlights the importance of keeping data accurate and up-to-date to fulfill the specified purposes and the need to process the data with due regard to confidentiality, using appropriate safeguards (organizational, administrative, physical, and technical) to protect the security of personal data.

To realize the desired and/or legislated standards of data protection, there are several concrete design and administrative arrangements that social protection programs can implement. Box 4.7 describes two such examples. First, the case of Turkey offers an example of an established social registry with good data privacy and data protection practices. The registry was developed a decade ago, before the adoption of the General Data Protection Regulation by the European Union. It was also before the recent debate about data privacy in the digital age, including the risk associated with automatic authentication through facial recognition (without the explicit and meaningful consent of the person being identified). The second example, from Morocco, presents a recent, holistic case of development of a data privacy and data protection framework for a twin unique ID registry and social registry.

Examples of How Countries Are Providing Data Security

Turkey has taken several steps to secure the data in its Integrated Social Assistance System (ISAS) (Ortakaya 2020). Security is particularly important because the system contains both personal profiles and financial information for more than 37 million people, but it is challenging because to carry out its function, ISAS is accessed by approximately 11,000 personnel in different functions (Ortakaya 2020). Two major risks associated with a management information system are corruption and the protection of privacy. Access to the data is highly controlled to prevent any tampering or leakage.

- ISAS employs a two-factor authentication process. Users are given a token that generates a one-time password that is required for entry into the system.
- Each user is given access to a different part of the system based on his/her specific roles and responsibilities. This system prevents unauthorized users from gaining access and allows the Ministry of Family and Social Policies to monitor usage.
- All queries made in the system are recorded with a barcode, which indicates the information that was queried, by whom, and on what date. Transactions made on the system are logged in a database and monitored.
- Institution staff who are responsible for providing database updates to ISAS are given access to the system via the virtual private network. Permission for this access is tied to the staff member's computer.
- The system's core hardware is also protected with security measures, and system rooms are monitored by cameras and sensors. Only authorized staff can enter the system rooms, and they can do so only by using an electronic card and fingerprint verification.
- Data flow within the system is encrypted according to international standards.

The Moroccan social registry and its underlying National Population Registry offers a more recent example of a data privacy and protection ecosystem encompassing legal and technical safeguards. The collection and use of information in the two registries are, first, subject to the national data privacy law and the Data Privacy Commission. In addition, specific legislation[a] and regulations underlying the two registries include additional data privacy provisions, which regulate the type of

> **BOX 4.7** *(continued)*
>
> information that can be collected, its uses, the rights of the data "own-ers," and procedural elements—such as delays for applicants to allow them to exercise their rights to review, contest, or update their records. Finally, the technological choices for the two registries incorporate sev-eral privacy-by-design principles in the way they handle personal data:
>
> - Data minimization (limited information capture for the purpose of establishing identity, anonymized numbering scheme, and mini-mal transaction records)
> - User consent and control (consent-based enrollment and authen-tication, extensive controls over use of identifications (IDs), and data portability)
> - Data security (encryption of data in motion and at rest and secure offline authentication)
> - Transparency (notification and real-time awareness of usage with untamperable data and secure and transparent audit records).
>
> a. See Law 18-72 on the targeting system for social assistance programs and the creation of the National Registries Agency (relative au dispositif de ciblage des bénéficiaires des programmes d'appui social et portant création de l'Agence nationale des registres), published in the Official Gazette 6950, August 8, 2020.

Conclusion

Delivery systems matter for targeting outcomes in normal times and in emergencies. This chapter provided a brief treatment of the topic as an overlapping team has recently produced an extensive companion volume on delivery systems (Lindert et al. 2020) and there are many materials already written or in the works on IDs, social registries, payment mecha-nisms, and adaptive social protection. Analysis and implementation of social protection programs should take these issues into account and con-sider how the delivery systems contribute to or solve problems in targeting. There is no phase of the delivery cycle that does not affect targeting outcomes, but the following are particularly common or important places to focus efforts to upgrade systems:

- Improving outreach and communication so that people who are meant to be served by programs are aware of them and know how to access them.
- Ensuring low transaction costs (in the time, travel, and mental band-width of those in pursuit of benefits and in calendar time in queue) and improving the client experience of inclusion and dignity.

- Developing dynamic intake processes so that all who are eligible can apply promptly rather than waiting years for the chance.
- Developing routine or ongoing recertification and exit processes with a periodicity to match the program objectives and expected dynamics of changes in households' welfare.
- Preparing in advance for expectable disasters and crises, with triggers and emergency rules of operation laid out.
- Building the client interface systems and capacities to run the programs well, with good governance and convenience for clients.
- Upgrading practices for data management and data protection apace with the greater use of technology in delivery systems.

Notes

1. https://socialprotection-humanrights.org/framework/.
2. The program has recently been transformed and no longer exists as a conditional cash transfer due to shifts in the social policy in Mexico.
3. Quilombolas communities are the current inhabitants of quilombos who were organized by fugitive African-American slaves in remote and hard-to-find areas. Nowadays, the quilombolas population is mostly composed of descendants who lived, in their majority, from subsistence agriculture on donated, bought, or long-occupied lands. According to the Commisão Pro-Índio de São Paulo, in 2019, there were 3,386 quilombolas communities across the country, but only 181 had land titles, while 1,719 were pursuing the process of acquiring legal title. For more on this, see Gaspar (2009).
4. Some questions are related to physical characteristics of the household; characteristics of household members, such as education level, employment status of the household head, and/or working-age adults; income information for all household members; description of the household's assets; and some basic household expenses (for example, rent and utilities).
5. https://www.apc.org/en/news/extreme-poverty-and-digital-welfare-new-report-un-special-rapporteur-extreme-poverty-raises.
6. Diagnostics on reducing barriers to inclusion in ID systems may include dimensions such as whether there are women-only registration units; mobile or door-to-door services; outreach and information campaigns; forms in local languages or braille; multilingual personnel or staff trained to assist disabled or illiterate groups; allowing nonbinary gender categories and procedures for changing gender attributes; and alternative procedures for those who are unable to provide biometrics, proof of citizenship, or other supporting documents for enrollment/authentication (World Bank 2018a).
7. Foundational ID systems (such as civil registries, national IDs, population registers, and so forth) are created to serve as authoritative sources of legal identity information for the general population and to provide proof of identity for a variety of public and private sector use cases. See World Bank (2019a, xiv).

8. Estimations based on data from the Global ID4D Dataset 2018, accessed from https://datacatalog.worldbank.org/dataset/identification-development -global-dataset.

9. Know-your-customer refers to the process of verifying the identity of applicants, either at the time of application or during cross-verification checks.

10. https://www.worldbank.org/en/news/feature/2016/02/04/pakistan-building -equality-for-women-on-a-foundation-of-identity.

11. During the COVID-19 crisis, the usual distinction between local and nonlocal *hukou* workers was temporary relaxed, and *dibao* eligibility was expanded to unemployed, low-income migrant workers.

12. The family certificate is an official document issued in several countries (Algeria, Belgium, China, France, Germany, Japan, the Republic of Korea, Morocco, Switzerland, and Vietnam). It consists of a collection of extracts from civil status documents relating to a family (births, deaths, marriages, and divorces). This type of document could be used by social programs as proof of family composition. The same objective can be achieved in countries that operate digital civil registries. The digital civil registry could identify the family unit—parents and minor children—based on the information from birth and death certificates. See https://www.service-public.fr/particuliers/vosdroits/N31784.

13. See "Fraud and Error in the Benefit System, 2017/18 Estimates," Department of Work and Pensions, www.dwp.gov.uk.

14. As an example, in Sierra Leone, in setting up the first cash transfer program in 2014–16, several steps were taken to reduce errors in this stage, among them: (1) financial literacy was assessed and enrollment information related to the payment system was designed in graphical form to facilitate use of the payment system; and (2) to keep both administrative costs for the program and travel costs for participants reasonable, communities were grouped into clusters with enrollment carried out in the lead community of the cluster.

15. Fraud refers to intentional behavior on the part of the benefit claimant to obtain a benefit to which she/he is not entitled, or a larger one. Error refers to unintentional mistakes on behalf of benefit claimants or staff in the benefit office. When the error is made by the claimant, it is called a customer error; if it is made by program staff, it is called an official error.

16. https://www.calpnetwork.org/wp-content/uploads/2020/01/phl-ctp-final.pdf; http://documents1.worldbank.org/curated/en/967551504637043989/pdf /Typhoon-Yolanda-Haiyan-and-the-case-for-building-an-emergency-cash -transfer-program-in-the-Philippines.pdf.

17. See Oseni et al. (2021) for a consolidated guidebook.

18. For example, in Colombia, in the municipal offices involved with the social registry, one-third of the staff have been in their positions only one or two years.

19. https://uidai.gov.in/ecosystem/training-testing-certification-ecosystem.html.

20. To get into the program, applicants may underreport or misreport some of their circumstances, if these are difficult or impossible to verify.

21. Data privacy is about the proper handling of data—how it is collected, stored, and used—and maintaining compliance with agreements and consent.

22. Data protection refers to who gets access to data and protecting it from unauthorized users through encryption, key management, and authentication.

23. https://www.unsystem.org/personal-data-protection-and-privacy-principles.

References

Alston, Philip. 2019. "Report of the Special Rapporteur to the UN on Extreme Poverty and Human Rights." A/74/48037, United Nations, New York.

Amjad, Rashid, M. Irfan, and G. M. Arif. 2015. "An Analysis of the Remittances Market in Pakistan." In *Pakistan: Moving the Economy Forward*, edited by Rashid Amjad and Shahid Javed Burki, 280–310. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781316271711.012.

ANPIS/ANPOFM (Ministry of Labor and Social Policy/National Agency for Payments and Social Inspection). 2021. "National Campaign for the Verification of the Eligibility for the GMI Program and the Identification of Measures to Increase the Employment Rate of Work-Able Beneficiaries." ANPIS/ANPOFM, Bucharest, Romania. https://www.mmanpis.ro/wp-content/uploads/2021/04 /Raport-privind-Campania-nationala-de-verificare-a-stabilirii-si-acordarii -ajutorului-social-si-pentru-identificarea-masurilor-de-crestere-a-ocuparii -persoanelor-apte-de-munca.pdf.

Azevedo, João Pedro. 2017. "Assessment of Social Benefit Effectiveness: An Analysis at the Municipal and City Level of Croatia." World Bank, Washington, DC.

Barca, Valentina, and Richard Chirchir. 2014. "Single Registries and Integrated MISs: De-Mystifying Data and Information Management Concepts." Department for Foreign Affairs and Trade, Government of Australia. https:// www.opml.co.uk/files/2018-05/barca-chirchir-2014-data-information -management-social-protection.pdf?noredirect=1.

Barca, Valentina, and Richard Chirchir. 2017. "Integrating Data and Information Management for Social Protection: Social Registries and Integrated Beneficiary Registries." Department for Foreign Affairs and Trade, Government of Australia. https://www.dfat.gov.au/sites/default/files/integrating-data-information -management-social-protection-full.pdf.

Bashir, Sajitha, Carl J. Dahlman, Naoto Kanehira, and Klaus Tilmes. 2021. *The Converging Technology Revolution and Human Capital: Potential and Implications for South Asia*. South Asia Development Forum. Washington, DC: World Bank. http://hdl.handle.net/10986/36156.

Beazley, Rodolfo, Ana Solórzano, and Valentina Barca. 2019. "Study on Shock-Responsive Social Protection in Latin America and the Caribbean: Summary of Key Findings and Policy Recommendations." Oxford Policy Management in collaboration with the World Food Programme, Oxford, UK. https://www .opml.co.uk/files/Publications/a1537-shock-responsive-social-protection-latin -america-caribbean/summary-of-key-findings-and-policy-recommendations .pdf?noredirect=1.

Boko, Joachim, Luis Alberro, Maimouna Gueye, Audrey Ariss, and Tina George. 2020. "Technology & Innovations for Africa's Human Capital in the Face of the COVID-19 Crisis: Technical Note on Togo's Novissi Program." World Bank, Washington, DC.

Bowen, Thomas. 2016. "Typhoon Yolanda (Haiyan) and the Case for Building an 'Emergency Cash Transfer' Program in the Philippines." Philippine Social Protection Note 10, World Bank, Manila, Philippines. http://documents1.worldbank.org /curated/en/967551504637043989/pdf/Typhoon-Yolanda-Haiyan-and-the-case -for-building-an-emergency-cash-transfer-program-in-the-Philippines.pdf.

Bowen, Thomas, Carlo del Ninno, Colin Andrews, Sarah Coll-Black, Ugo Gentilini, Kelly Johnson, Yasuhiro Kawasoe, Adea Kryeziu, Barry Maher, and Asha Williams. 2020. *Adaptive Social Protection: Building Resilience to Shocks*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/33785.

Cho, Yoonyoung, Jorge Avalos, Yasuhiro Kawasoe, Doug Johnson, and Ruth Rodriguez. 2021. "Mitigating the Impact of COVID-19 on the Welfare of Low Income Households in the Philippines: The Role of Social Protection." COVID-19 Low Income HOPE Survey Note No. 1, World Bank and Australia Department of Foreign Aid and Trade, Washington, DC. https://reliefweb.int/sites/reliefweb.int/files/resources/Mitigating-the-Impact-of-COVID-19-on-the-Welfare-of-Low-Income-Households-in-the-Philippines-The-Role-of-Social-Protection.pdf.

Cho, Yoonyoung, Yasuhiro Kawasoe, Ruth Reyes Rodriguez, and Myra Valenzuela. 2021. "COVID-19 G2P Cash-Transfer Payments: Case Study: Philippines." World Bank, Washington, DC. http://documents.worldbank.org/curated/en/368351625163110870/COVID-19-G2P-Cash-Transfer-Payments-Case-Study-Philippines.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. "Targeting of Transfers in Developing Countries: Review of Lessons and Experience." World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/14902.

Deryugina, Tatyan. 2016. "The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance." Working Paper 22272, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w22272.

Devereaux, Stephen, Edoardo Masset, Rachel Sabates-Wheeler, Michael Samson, Althea-Maria Rivas, and Dolf te Lintelo. 2017. "The Targeting Effectiveness of Social Transfers." *Journal of Development Effectiveness* 9: 2: 162–211. https://doi.org/10.1080/19439342.2017.1305981.

DNP (Departamento Nacional De Planeación). 2016. "CONPES 3877: Declaración De Importancia Estratégica Del Sistema De Identificación De Potenciales Beneficiarios (Sisbén IV)." DNP, Bogotá, Colombia. https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3877.pdf.

Enabling Digital & GIZ. 2020. "Implementation Guide: Good Practices for Ensuring Data Protection and Privacy in Social Protection Systems, A Guide for Practitioners Advising and Working in Low- and Middle-Income Countries." SPIAC-B Workstream on Data Protection (Working Group on Digital Social Protection), Enabling Digital & GIZ, Bonn, Germany.

Gaspar, Lúcia. 2009. "Quilombolas." Pesquisa Escolar Online, Joaquim Nabuco Foundation, Recife, Brazil.

Gelb, Alan, and Anna Diofasi Metz. 2018. *Identification Revolution: Can Digital ID Be Harnessed for Development*? Washington, DC: Center for Global Development.

Gelb, Alan, and Anit Mukherjee. 2020. "Digital Technology in Social Assistance Transfers for COVID-19 Relief: Lessons from Selected Cases." Policy Paper 181, Center for Global Development, Washington, DC. https://www.cgdev.org/publication/digital-technology-social-assistance-transfers-covid-19-relief-lessons-selected-cases.

Gentilini, Ugo, Mohamed Almenfi, Ian Orton, and Pamela Dale. 2020. "Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures." Living Paper. World Bank, Washington, DC. http://hdl.handle.net/10986/33635.

GSMA (Global System for Mobile Communication Association). 2020. *Connected Women: The Mobile Gender Gap Report 2020*. London: GSMA. https://www.gsma .com/mobilefordevelopment/wp-content/uploads/2020/05/GSMA-The -Mobile-Gender-Gap-Report-2020.pdf.

Hidrobo, Meliisa, Lieven Huybregts, Naureen Karachiwalla, and Shalini Roy. 2020. "Multi-Component Cash Transfer Programs: Evidence from Mali's Social Safety Net Program (Jigisémèjiri)." Project Note. International Food Policy Research Institute, Washington, DC. https://doi.org/10.2499/p15738coll2.133601.

Hidrobo, Melissa, Shalini Roy, Malek Abu-Jawdeh, Lieven Huybregts, Leila Njee-Bugha, Eric Sessou, and Yves Kameli. 2015. "Filets Sociaux (Jigisémèjiri)." Program Baseline Report. International Food Policy Research Institute, Washington, DC.

Hidrobo, Melissa, Shalini Roy, Lieven Huybregts, Leila Njee-Bugha, Eric Sessou, and Yves Kameli. 2018. "Filets Sociaux (Jigisemejiri)." Program Midline Report. International Food Policy Research Institute, Washington, DC.

Hill, Ruth, and Catherine Porter. 2017. "Vulnerability to Drought and Food Price Shocks: Evidence from Ethiopia." *World Development* 96. https://www .sciencedirect.com/science/article/pii/S0305750X15306392.

Isik-Dikmelik, Aylin. 2012. "Do Social Benefits Respond to Crises? Evidence from Europe and Central Asia during the Global Crisis." Social Protection & Labor Discussion Paper 1219, World Bank, Washington, DC. https://openknowledge .worldbank.org/handle/10986/13565.

ITU (International Telecommunication Union). 2020. ITU, Geneva (accessed August 2, 2020). https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

Leite, Phillippe, Tina George, Changqing Sun, Theresa Jones, and Kathy Lindert. 2017. "Social Registries for Social Assistance and Beyond: A Guidance Note and Assessment Tool." Social Protection & Labor Discussion Paper 1704, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle /10986/28284.

Lindert, Kathy. 2017. "Communications & Outreach in Social Safety Nets." Presentation in the Safety Nets and Delivery System Core Course, March, World Bank, Washington, DC.

Lindert, Kathy, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Nishikawa Chávez. 2020. *Sourcebook on the Foundations of Social Protection Delivery Systems*. Washington, DC: World Bank. http://hdl.handle.net/10986/34044.

Medellín, Nadin, Pablo Ibarrarán, Marco Stampini, and Juan Miguel Villa. 2015. "Moving Ahead: Recertification and Exit Strategies in Conditional Cash Transfer Programs." Inter-American Development Bank, Washington, DC. https://publications.iadb.org/en/moving-ahead-recertification-and-exit -strategies-conditional-cash-transfer-programs.

Mori, Hideki, Yuka Makino, Kenichi Nishkawa Chavez, Rajib Shaw, Adea Kryeziu, Raiden Dillard, Thomas Vaughan Bowen, Barry Patrick Maher, Matthew Liam Hobson, Ana-Veronica Lopez, Erika Vargas, Phillippe Leite, Semeon Joachim Abel, Rich Webster, Galina Kalvatcheva, and Llanco Talamantes. 2020. *Social Protection and Labor – Disaster Risk Management Toolkit*. Washington, DC: World Bank Group. http://documents.worldbank .org/curated/en/758321601534340630/Social-Protection-and-Labor -Disaster-Risk-Management-Toolkit.

Muller, Katharina. 2016. "The Political Economy of Universal Pensions in Bolivia." *Social Policy* 1: 1–26. https://www.sozialpolitik.ch/fileadmin/user_upload/The -Political-Economy-of-Universal-Pensions-in-Bolivia.pdf.

Nishtar, Sania. 2020. "Ehsaas Emergency Cash: A Digital Solution to Protect the Vulnerable in Pakistan during the COVID-19 Crisis." Government of Pakistan. https://www.pass.gov.pk/Document/Downloads/Ehsaas%20Emergency%20 Cash%20Report%20Oct%202020_Dec15_2020.pdf.

OECD (Organisation for Economic Co-operation and Development). 2020. *Countering Fraud in Social Benefit Programmes: Taking Stock of Current Measures and Future Directions*. Paris: OECD Publishing. https://www.oecd.org/gov /countering-fraud-in-social-benefit-programmes-71df2657-en.htm.

Ohlenburg, Tim. 2020. *AI in Social Protection – Exploring Opportunities and Mitigating Risks*. Bonn: GIZ. https://www.adb.org/publications/ai-social-protection-exploring -opportunities-mitigating-risks.

OPM (Oxford Policy Management). 2017. *Shock-Responsive Social Protection Systems Research: Literature Review*, 2nd edition. Oxford, UK: OPM. https://www.opml .co.uk/files/Publications/a0408-shock-responsive-social-protection-systems /srsp-literature-review.pdf?noredirect=1.

Ortakaya, Ahmet Fatih. 2020. "COVID-19 Social Support Program in Turkey." World Bank, Washington, DC.

Oseni, Gbemisola, Amparo Palacios-Lopez, Harriet Kasidi Mugera, and Josefine Durazo. 2021. *Capturing What Matters: Essential Guidelines for Designing Household Surveys*. LSMS Guidebook. Washington, DC: World Bank Group. http:// documents.worldbank.org/curated/en/963531615786610706/Capturing- What-Matters-Essential-Guidelines-for-Designing-Household-Surveys.

Palacios, Robert. 2020. "Scaling Up Social Assistance Payments as Part of the COVID-19 Pandemic Response." G2Px. World Bank, Washington, DC. http:// pubdocs.worldbank.org/en/655201595885830480/WB-G2Px-Scaling-up-Social -Assistance-Payments-as-Part-of-the-Covid-19-Pandemic-Response.pdf.

Pelly, Isabelle, David de Wild, and Camila Inarra. 2015. "Philippines Haiyan Response – A Multi-Sectoral Review of the Use of Market Analysis and the Design and Implementation of CTPs." Save the Children, London. https:// www.calpnetwork.org/wp-content/uploads/2020/01/phl-ctp-final.pdf.

Porter, Catherine, and Emily White. 2017. "The Development of Vulnerability Functions Relating Household Poverty Outcomes to Crop Failures in Ethiopia with the Prospect of Developing a Probabilistic Catastrophe Risk Model." *The Geneva Papers on Risk and Insurance: Issues and Practice* 42 (4): 657–74. https:// econpapers.repec.org/article/palgpprii/v_3a42_3ay_3a2017_3ai_3a4_3ad_3a1 0.1057_5fs41288-017-0051-6.htm.

Rofman, Rafael, and Maria Laura Oliveri. 2012. "Pension Coverage in Latin America: Trends and Determinants." Social Protection & Labor Discussion Paper 1217, World Bank, Washington, DC. https://openknowledge.worldbank .org/handle/10986/13561.

Sepulveda Carmona, Magdalena. 2018. "Is Biometric Technology in Social Protection Programmes Illegal or Arbitrary? An Analysis of Privacy and Data Protection." Extension of Social Security Working Paper 59, International Labor Organization, Geneva. https://www.researchgate.net/publication /325909014_Is_biometric_technology_in_social_protection_programmes_illegal _or_arbitrary_An_analysis_of_privacy_and_data_protection.

Silva, Veronica; Heidi Berner, Luis Díaz Silva, and Javiera Troncoso. 2018. "Registro Social de Hogares de Chile." http://www.desarrollosocialyfamilia.gob.cl/storage/docs/RSH_paper_2.pdf.

Sluchynskyy, Oleksiy. 2019. "Better Data for Better Social Programs and Policies Enabled by e-Government Solutions: Case Studies." World Bank, Washington, DC. https://thedocs.worldbank.org/en/doc/674741548878149565-0160022019/original/10.45amJan15DeliverySystemSluchynskyInteroperabilityanddatasharing.pdf.

SPACE (Social Protection Approaches to COVID-19: Expert Advice). 2021. "Useful COVID-19 and Social Protection Materials." https://socialprotection.org/sites/default/files/publications_files/SPACE%20Useful%20COVID-19%20and%20Social%20Protection%20Materials%20V3%20%281%29.pdf.

Tesliuc, Emil, Lucian Pop, Margaret Grosh, and Ruslan Yemtsov. 2014. *Income Support for the Poorest: A Review of Experience in Eastern Europe and Central Asia*. Directions in Development: Human Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/18886.

Thinkthrough Consulting. 2021. "Analytical Deep Dive on SSA Registration Procedure in Nepal." Thinkthrough Consulting, New Delhi, India.

UNECA (United Nations Economic Commission for Africa). 2021. "Civil Registration Systems as Enablers of Emergency Response to the COVID-19 Crisis: Namibian Emergency Income Grant." Technical Brief #2, UNECA, Addis Ababa, Ethiopia. http://www.apai-crvs.org/sites/default/files/public/Technical%20brief%20no.2.pdf.

UNICEF (United Nations Children's Fund). 2019. *Programme Guidance: Strengthening Shock Responsive Social Protection Systems*. New York: UNICEF. https://www.unicef.org/documents/programme-guidance-strengthening-shock-responsive-social-protection-systems.

United Kingdom National Audit Office. 2006. *International Benchmark of Fraud and Error in Social Security Systems*. London: United Kingdom National Audit Office. https://www.nao.org.uk/wp-content/uploads/2006/07/05061387.pdf.

UTGFS (Unité Technique de Gestion Filets Sociaux). 2014a. "Enquête de Suivi Post-Transfert." UTGFS, Bamako, Mali.

UTGFS (Unité Technique de Gestion Filets Sociaux). 2014b. "Evaluation de la phase pilote du processus de ciblage et des transferts monétaires." UTGFS, Bamako, Mali.

UTGFS (Unité Technique de Gestion Filets Sociaux). 2014c. "Mission de Contrôle du Processus de Ciblage dans les Cercles de Kolokani et Dièma - Période du 15 au 25 Septembre 2014." UTGFS, Bamako, Mali.

UTGFS (Unité Technique de Gestion Filets Sociaux). 2014d. "Rapport de Mission de Contrôle du Processus d'Enregistrement et de Collecte des données des bénéficiaires dans la région de Sikasso, Période du 23 au 28 Novembre 2014." UTGFS, Bamako, Mali.

Van Stolk, Christian, and Emil D. Tesliuc. 2010. "Toolkit on Tackling Error, Fraud and Corruption in Social Protection Programs." Working Paper 764, Rand Corporation, Santa Monica, CA. https://www.rand.org/pubs/working_papers/WR746.html.

Velarde, Rashiel. 2018. "The Philippines' Targeting System for the Poor: Successes, Lessons and Ways Forward." Social Protection Australian Aid Policy Note 16, World Bank, Washington, DC. http://hdl.handle.net/10986/30972.

World Bank. 2018a. "Global ID Coverage, Barriers, and Use by the Numbers: Insights from The ID4D-Findex Survey." World Bank, Washington, DC. https://documents1.worldbank.org/curated/en/953621531854471275/pdf/Global-ID-Coverage-Barriers-and-Use-by-the-Numbers-Insights-from-the-ID4D-Findex-Survey.pdf.

World Bank. 2018b. *Guidelines for ID4D Diagnostics*. Washington, DC: World Bank. https://documents1.worldbank.org/curated/en/370121518449921710/Guidelines-for-ID4D-Diagnostics.pdf.

World Bank. 2018c. "Moldova: Strengthening the Effectiveness of the Social Safety Net Results-Based Financing Specific Investment Loan Project." World Bank Group, Washington, DC. https://imagebank2.worldbank.org/Search/30251568.

World Bank. 2019a. *ID4D Practitioner's Guide*. Washington, DC: World Bank. https://id4d.worldbank.org/guide.

World Bank. 2019b. "Restructuring Paper: Argentina: Children and Youth." Project Report No. RES38682, World Bank, Washington, DC.

World Bank. 2020a. "Accelerating India's Covid-19 Social Protection Response Program." Report No. 147337-IN, World Bank, Washington, DC.

World Bank. 2020b. "Brazil: Income Support for the Poor Affected by COVID-19 Project." World Bank, Washington, DC.

World Bank. 2020c. "Colombia: Support to the Development of Colombia's Integrated Social Registry." Report Number PAD4202, World Bank, Washington, DC.

World Bank. 2020d. "Emergency Locust Response Program." Report Number PAD 3904, World Bank, Washington, DC.

World Bank. 2020e. "Ethiopia: Strengthen Ethiopia's Adaptive Safety Net Project." World Bank Group, Washington, DC.

World Bank. 2020f. "Unraveling Data's Gordian Knot: Enablers & Safeguards for Trusted Data Sharing in the New Economy." World Bank, Washington, DC. http://hdl.handle.net/10986/35119.

World Bank. 2021. "Stress Testing Social Protection: A Guide for Practitioners Version 1." World Bank, Washington, DC. https://documents1.worldbank.org/curated/en/559321634917529231/pdf/Stress-Testing-Social-Protection-A-Rapid-Appraisal-of-the-Adaptability-of-Social-Protection-Systems-and-Their-Readiness-to-Scale-Up-A-Guide-for-Practitioners.pdf.

World Bank and DFAT (World Bank and Department of Foreign Affairs and Trade). Forthcoming. *Diverse Paths: The Dynamic Evolution of Social Protection in Asia and the Pacific*. Washington, DC, and Sydney, Australia: World Bank and Australian Department of Foreign Affairs and Trade.

World Bank and FCDO (UK Foreign, Commonwealth, and Development Office). 2021. "Verificação dos Critérios de Elegibilidades, Seleção de Beneficiciários e Cruzamentos de Bancos De Dados no Âmbito do Auxílio." Technical Note, World Bank and FCDO, Brasilia, Brazil.

World Bank Group. 2016. *World Development Report 2016: Digital Dividends*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/23347.

5

Choosing among Targeting Methods

Margaret Grosh, Phillippe Leite, Matthew Wai-Poi,
Emil Tesliuc, Juul Pinxten, Nina Rosas,
and Priyanka Kanth

Introduction

The choice among different targeting methods must be grounded in under-standing the larger hierarchy between policy objectives, program design, targeting methods, implementation, and metrics for measurement. A pol-icy objective is an overarching goal, such as "learning for all" or "reduction of poverty." A program is an intervention that is implemented to achieve the policy objective and can be population wide or for a certain subset of the population. The program design refers to all the parameters for the pro-gram—who it is meant to serve, the benefits and services to be provided, the duration of these, and so forth. A targeting method is a tool to identify the population intended to be served by a specific program, to conduct eli-gibility assessments. Implementation affects all elements of program design. This book is particularly concerned with the elements around eligibility determination. Targeting metrics show how well the program reached the intended population and the associated costs. A fuller impact evaluation helps to discern how the program changed key outcomes such as poverty, inequality, participation in the labor force and earnings, savings or invest-ment in enterprises, the use of education or health services, and any of a long list of education or health outcomes. Figure 5.1 provides a simplified

Figure 5.1   Hierarchy of Action

EXAMPLES

Policy objectives
- Poverty reduction
- Education for all
- Female empowerment
- Universal social protection

Program level
- Provide a specific set of services or benefits to a defined intended population, for example, income support to the poor, scholarships to out-of-school youth, or reproductive health services to adolescents
- Generally, multiple programs help to achieve a single goal
- Similarly, single programs may contribute to multiple goals

Delivery system
- Outreach
- Assessment of needs and conditions
- Enrollment
- Provision of benefits or services

Targeting method
- Self-targeting
- Geographic
- Demographic
- Welfare based
  - Means test
  - Hybrid means test
  - Proxy means test
  - Community-based targeting

Targeting metrics
- Dichotomous, for example, errors of inclusion, errors of exclusion, and targeting differential
- Continuous, for example, coverage, incidence, and distributional characteristic
- Costs, for example, administrative costs, transaction costs, stigma or social discord, and incentives

Impacts
- On poverty and distribution
- On human capital
- On labor activity

Source: Original compilation for this publication.

overview of these elements. Reality is more complex as there are interactions along the hierarchy—especially between the choice of targeting method and the program's delivery system.

To discern the best way to achieve improvements requires understanding where along the hierarchy problems occur. For example, if a high level of errors of exclusion is observed, it is important to sort out whether that is because the program's budget and decisions about generosity lead to a

smaller number people who can be served than the size of the intended population, or whether the intended population does not know about the program and apply for it, or whether eligibility assessments are incorrect in a significant number of cases.

To achieve high-level social policy objectives, countries offer a myriad of programs to their residents, including some that are universally accessible and some that are more narrowly focused on a particular group. Social protection programs often have multiple outcomes and can cater to multiple policy objectives; similarly, a policy objective could be supported by several different programs often directed at specific population groups. Social protection contributes to the high-level goals of achieving reductions in poverty and inequality, handling risk, building human capital and prosperity, and mitigating and adapting to climate change. As chapter 1 explains, the policy objective of universal social protection is supported by social assistance and social insurance programs of many designs and intentions. The programs include guaranteed minimum income and other unconditional and conditional cash transfers; child allowances and social pensions; school feeding; food stamps and heating assistance; productive inclusion and training; unemployment, disability, and maternity benefits; and pensions plans.

Policy makers determine the population of focus for each social program based on high-level objectives, followed by an analysis of desired outcomes and the patterns of gaps or differences in those outcomes across the population, subject to available resources and notions of the social contract/political economy of the country. For example, to help meet an antipoverty objective, some countries may aim to extend income support to all poor people and choose to introduce a poverty-targeted program that covers families of any composition; other countries might start with a social pension to help protect the elderly from poverty. To help meet a societal goal that all children by age five are well prepared for school, countries will carry out a diagnostic. Some may find that the problem is inadequate quality or coverage of services for health, education, and stimulation programs for preschool children. Others may find that the problem is less rooted in the supply of services and more in the ability of poor households to provide food, shelter, and supportive parental attention, and that extra support for these families is needed through cash or in-kind assistance. Once the population of focus for a given program is loosely determined, policy makers need to work through a series of issues, such as those handled in chapter 3, to make more precise whom they seek to serve.

While the list of targeting methods is not new[1] (see table 5.1), the lessons and experience with them have expanded markedly in recent years—how they can be adapted to or combined in different contexts; how they have evolved with growing social protection programming, capacity, and ambition; and with the data revolution. This chapter sketches the list of

Table 5.1   Common Targeting Methods

All programs have an element of self-selection:
Self-targeting. *Implicit.* All programs are implicitly self-targeted in that individuals decide to participate or apply if they consider the package of benefits and program rules acceptable. As discussed in chapter 2, many elements of program implementation that affect transaction costs and stigma affect that calculation and degree of self-targeting. *Explicit.* Some programs also have explicit design features to promote differential take-up across who the population. Public works programs pay low wages for short periods, with the idea that those with better employment will not participate. Food subsidies or rations may feature nutritious staple foods that are a larger share of the diets of the poor than the less poor, and sometimes the less prestigious versions—broken rice, coarse flours, unattractive packaging.
Some programs operate by defining broad categories of eligible households:
Geographic targeting. In the strict version, a program selects/serves only those in a defined geographic area; in a more moderate version, a program allocates rationed caseloads to different areas based on spatial variation in need. Either version requires the ability to delineate boundaries, which may be clear for units of political representation (state, district) but less so for smaller areas (village, neighborhood). And it may entail some requirement of "belonging" or duration or formal registration of residency.
Demographic/categorical targeting. This applies when benefits are granted to people according to their membership in fairly easy-to-observe categories. The categories that are most commonly used and easiest to observe are based on age, civil status, and gender, though programming may also be directed to veterans on a categorical basis. Ethnicity is occasionally used, as in affirmative action programs.
Some programs seek to distinguish the welfare of specific households:
Means testing (MT). When household's income and/or assets determine eligibility, often these are verified against independent sources.
Hybrid means testing (HMT). When a significant part of the information on the family or household's socioeconomic condition can be verified against independent sources, and the other part needs to be imputed or predicted.
Proxy means testing (PMT). When information on the family or household's socioeconomic condition needs to be estimated/predicted based on (mostly) observable sociodemographic characteristics and economic assets because verification of socioeconomic status cannot be performed.
Community-based targeting (CBT). When community leaders or members use information known to them from day-to-day living in the community to guide or choose who should be in or out of the program. As part of this assessment, the community may be guided to use wealth ranking or household economic analysis (HEA) techniques or similar techniques.
Some people use the term **affluence testing** to refer to programs with the eligibility threshold set quite high up the welfare distribution rather than toward the bottom end. Affluence testing is not a method per se in that it isn't about what information is used to determine needs. It is often associated with means testing but can apply to HMT and PMT as well.
Some programs seek to ration without further ranking or comparison of need:
Public lottery. When a random process is used to ration spaces among eligible applicants in an oversubscribed program. In a sense, this is less of a targeting method as such and more an additional way to ration selection.

Source: Original compilation for this publication.

methods, provides some observations on the factors that have influenced their development, and discusses elements that may help guide the selection of the method or methods in a given context. It also examines how frequently different methods are used and combined. The attention to delivery systems covered in chapter 4 is also a core topic for all the methods covered in this chapter, as a good delivery system is required for any method to work optimally. The choice of targeting method in a particular setting should also reflect a deep knowledge of the how-tos of the candidate method(s), a topic taken up in chapter 6. Thus, these two chapters might be read iteratively.

Patterns in Using and Combining Targeting Methods

Before turning to considerations for choosing among the different methods, this section briefly examines how commonly each method is used. The data come from a database that is broad but as yet imperfect for the purpose.[2] The Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE) database includes distributional performance indicators derived from household survey–based data on program categories, as described in chapter 2, as well as administrative data at the program level. The administrative data include information on program size (number of beneficiaries and amount of spending) and design parameters including, for social assistance programs, the targeting methods used. Data are available on the targeting methods used for 1,985 of the 2,623 social assistance and general subsidies programs included in the ASPIRE administrative database. The targeting method(s) used for each program was coded by a consultant or World Bank staff member who contributed that particular line of data to the global data set. It was possible to code multiple targeting methods for a program.

About two-thirds of the programs use a single targeting method, with the remaining third using mixed methods. The majority of the latter category combine only two methods (68 percent of those with multiple methods); the remainder (only 10 percent of all programs) use three or more methods. This may be a lower bound on the use of mixed methods. A light scan of the raw data identified several programs that might have coded multiple methods, as apparently sometimes the coders recorded only what they thought of as the most important method. Categorical (as recorded in this data set) and means testing are the main methods that are used by themselves. Geographic, proxy means testing (PMT), and community-based targeting (CBT) are rarely used alone.

Among the social assistance programs with information on the targeting method, about 40 percent use a household-level targeting method. This section counts programs that use either means testing/HMT),[3] PMT, CBT,

or a combination. Among the programs that use household targeting, 62 percent are coded as using a means test, 31 percent as using PMT, and 22 percent as using CBT. The percentages sum to more than 100 because programs can use multiple methods; for example, this book discusses several examples of programs that combine PMT and CBT. Of the programs that use household-specific methods, 40 percent use only one method, and most often it is means testing. Where a household-specific method is combined with another method, the overwhelming complementary method is categorical (such as a means-tested child allowance or social pension, or a guaranteed minimum income with different filters for different categories of individuals).

Among the programs that were observed, three-quarters used a categorical method other than geographic targeting, either stand-alone or with another method. Many programs are inherently categorical in design. In looking at the types of programs coded as using categorical targeting, there is a predominance of those related to age, such as family and child allowances, nutrition programs, school feeding, scholarships, provision of school supplies, old-age pensions, and burial grants. These accounted for about three-quarters of all those labeled "categorical targeting" in this data set. The next largest set of categorically targeted programs are disability programs, about 10 percent. Programs for war veterans account for another 5 percent. Categorical targeting is also used in 43 percent of programs that are not inherently categorical, such as poverty alleviation programs, targeted subsidies, or emergency support.

Geographic targeting, which is conceptually another sort of categorical targeting but fortunately coded separately in these data, is used in a quarter of the programs but rarely as the only method. Again, this is interpreted as a lower bound figure due to undercoding of multiple methods. Some programs in Africa that are less than national in scope did not include geographic targeting in their coding, although in the framework of this book, they are called geographically targeted programs. The book also adds "geographic" to the descriptors in more subtle cases where the program is national but rationed and so the caseload is allocated geographically, although the coders did not code them as geographic.

The choice of methods varies by region and income level in unsurprising ways. Low-income and lower-middle-income countries are more likely to use geographic targeting (23 and 17 percent of the programs in these countries, respectively). This method is used relatively infrequently in upper-middle-income and high-income countries (5 and 4 percent, respectively). Lower income countries are also most likely to use community-based methods (14 percent for low-income countries and 8 percent for lower-middle-income countries) compared with richer countries, which almost never use CBT and are more likely to use means testing (29 percent of

upper-middle-income programs and 23 percent of high-income ones). Perhaps mirroring income levels, Sub-Saharan African countries are most likely to use CBT and least likely to use means testing, which is particularly predominant in Eastern Europe and Central Asia. PMT is generally used in around 6–13 percent of the programs, less so in Sub-Saharan Africa, and infrequently in Eastern Europe and Central Asia.

The rest of the chapter looks at the different considerations in choosing which methods to use. The wide range of factors to be thought through in different contexts helps explain the diversity of use of methods.

Reflections on Patterns of Use of Targeting Methods

The tenor of the conversation around self-targeting seems to have changed over the years with respect to price subsidies. In the era of state-led development with major interventions in the agricultural markets in the form of import/export controls, state marketing boards, strategic grain reserves, and the like, general food price subsidies were fairly common and part of interventions with a mix of agricultural, social protection, and state-building objectives. The choice of food commodities to subsidize has become a less common thread of the social protection practice due to reduced interventions in markets and eventually due to the development of cash transfer programs. In 2008, during the general food and fuel price crisis, food subsidies or tax exemptions were a common response, especially in low-income countries that did not have well-established alternative social assistance–28 countries increased their subsidies and 84 reduced taxes on selected foodstuffs (IMF 2008). That is the case in fewer countries now, and although the responses to COVID-19 included food commodity distribution programs with various degrees of targeting, resorting to general food price subsidies and tax reductions has not been so common. In addition, there has been a secular move to cash rather than in-kind benefits over the years. For example, Indonesia's subsidized but low-quality rice program (variously named Implementation of Special Market Operation, Raskin, and then Rastra), which has been in place since the Asian Financial Crisis, was phased out in 2017 and replaced with a digital food voucher Food Assistance Program, Bantuan Pangan Non-Tunai (BPNT), then called Sembako (Holmemo et al. 2020; Banerjee et al. 2021).

Of course, every program is self-targeting in the sense that people have to deem the benefits worthy of the costs of participating. This is the flip side of the issues of transaction costs and stigma discussed in chapter 2. The growing prevalence of human rights viewpoints and their concern with inclusion and dignity rules out the always quite rare suggestion that delivery systems should be purposefully inconvenient. The accumulated work

on delivery systems should be improving inclusion, although there is still a journey to go. Thus, the main way in which people are purposefully induced to self-exclude due to transaction costs is in setting low wages for temporary employment on public works projects, which are discussed later in this chapter. Occasionally, self-exclusion from the top is encouraged for nearly universal programs, for example, India's Give It Up campaign, which was implemented as part of the liquid petroleum gas–related cash program, reaching 177 million people, and successfully promoted the self-exclusion of about 10 million wealthy individuals (Gelb and Mukherjee 2019).

In its traditional mode, geographic targeting was about rationing, locating programs exclusively, or concentrating caseloads in the neediest areas. Because this could be done with off-the-shelf data, it was much simpler than household-specific targeting, although of course, rationing carries explicit errors of exclusion.

As adaptive social protection gains prominence, especially in the contexts of natural disasters and climate change, geographic targeting has become a central method for preparing for and responding to such covariate shocks. Early warning systems, disaster risk management, and risk profiling analyses are helping countries to identify ex ante the needs of the population exposed to a shock and the areas that are more likely to be affected. This facilitates policies to help improve the resilience of communities and people. Better interoperability of an early warning system and a disaster risk management system can help improve preparedness and response to shocks.

As big data (for example, from satellite imagery, mobile phones, or digital content) has become more pervasive, the sophistication of information that can be brought to bear on geographic targeting has increased enormously,[4] but big data do not completely solve the problem of household-specific targeted income support programs. Big data make poverty mapping possible in places without recent censuses or surveys, with faster updates and more granularity than was previously possible. For example, poverty maps based on night lighting (sensing) may use a kilometer grid (see, for example, Skoufias, Strobl, and Tveit 2017). Even for poverty maps that can observe some dwelling-specific characteristics, such as the material of the roof, this is only a single characteristic and not often used alone in determining eligibility or setting benefit levels. The sole, although quite prominent, example for cash transfers[5] known to the authors is the Give Directly program, which initially used thatch roofs as its targeting criterion (Abelson, Varshney, and Sun 2014; Haushofer and Shapiro 2016) but eventually added additional criteria (Ohlenburg 2020). The *World Development Report 2021* (World Bank 2021b) envisages the use of mobile phone and social media data to target households directly, but several steps will be required to ensure access to these usually private sector data (see chapter 6 for

further discussion). More promising is the use of sensing data to trigger vertical or horizontal program expansions in response to droughts or floods.

Nonetheless, machine learning and big data approaches require rigorous benchmarking and assessment. The accuracy of big data poverty maps needs to be assessed; as chapter 6 discusses, much of their evaluation has been in-sample. Most importantly, machine learning does not replace the need to invest in people and human process, as Joshua Blumenstock (CEGA[6] Faculty Co-Director) highlighted in an interview[7]: "The algorithms are sort of the shiny object, and they receive a lot of attention. But when it comes to actually implementing social protections, going the last mile to put money in the hands of people who need it, the algorithms are just one small link in a much larger chain of humanitarian assistance. Most of the other links are human. Algorithms can help surface relevant information, but humans must decide what to do with it."

An important number of programs rely on demographic (also known as categorical age-based) targeting. Age (for children or the elderly) is used as the sole standard or combined with other criteria for eligibility for many programs, and benefit levels may be customized by age as well. There seem to be three variants in the reasoning to support demographic targeting, which are sometimes not clearly distinguished or acknowledged:

- First, some programs simply consider that all members of a group with a (usually) simple-to-observe characteristic deserve public support no matter their individual money-metric welfare or that of their families. Examples include veterans who provided service to their country and merit support/recognition for that. The blanket argument accords well with rights-based arguments as well, along the lines that children are inherently vulnerable and the precious future, society must nurture them, providing health, education, water, social protection, and so forth. Similarly, the elderly are vulnerable and deserving of support and respect for their service and wisdom.
- A second rationale acknowledges that not every member of a group requires public assistance within a money-metric notion of welfare but sees that, on the whole, members of the group and the families with which they live are poorer than average. In this line of argument, demographic targeting may have significant errors of inclusion, but the use of a single, easy-to-observe proxy is simpler to implement and more transparent than many other methods. Moreover, the groups selected resonate with societal views of deservingness in most places. Thus, it may be relatively easy to build consensus in support of such programs and any errors of inclusion tend not to offend.
- A third variant of demographic targeting is when programs or benefits that use a money-metric gradient for narrowly targeting only admit and provide for families that have members of the defined category. In the

United States, for example, the welfare state is both heavily means tested and markedly focused on children or the elderly. Many programs provide support to families with poor children or elderly members, but only to these families. Social assistance units composed of only poor, prime-age adults were long excluded altogether; even now, prime-age adults are required to meet much more stringent thresholds or work requirements, recertify more often, and so forth. There is enough consensus over societal responsibility for children and the elderly to support social protection programs in their favor, but there is much less consensus on supporting prime-age adults, especially if they are not working, except possibly for those living with serious disabilities.

Means testing is a common method used to differentiate eligibility and benefits, especially in highly formalized economies. Means testing is widely used in Western Europe and the long-standing Organisation for Economic Co-operation and Development members, such as Australia, Canada, Japan, New Zealand, and the United States, where verification of means is possible. Unverified means testing is used in some countries with substantial informality, such as Brazil and South Africa. Means testing is deemed something of a gold standard among household-specific methods because, unlike other methods, it contains no inherent measurement error, although of course, various errors creep in during implementation. Means testing has been shown to be very accurate in its assessments.

An allied strand of targeting practice grew from the European and Central Asian transition economies with the HMT. These countries had high but decreasing levels of formality and an orientation toward Western Europe with its high formality and use of verified means testing. The countries invented what Tesliuc et al. (2014) call an HMT that uses declared income that can be verified with an imputation or proxy for other sources of income that are not easily verifiable. The method has not spread as widely as PMT, but as the data revolution has increased the scope and decreased the cost of databases everywhere, it may be pertinent in places not using it now. HMT may be especially useful in moderately formal economies or for programs that use affluence tests, which try to screen out the top of the income distribution, which may have formal incomes or assets, more than trying to focus on the very poorest.

One of the strands of modern targeting practice in the developing world today is PMT, which originated in Latin America. Countries in that region had high income inequality, high levels of informality, relatively strong government and information, and largely adequate physical access to health and education services but big gaps in human capital outcomes. They also had years of fiscal, economic, and societal scarring from the debt crisis of the 1980s and all that ensued. Given the inequalities and limited fiscal space, household-level targeting was desired, but with high levels of informality, means testing as traditionally practiced in Western Europe and

North America seemed unreachable. Thus, countries such as Chile, Colombia, and Mexico started using PMT for a variety of programs, their conditional cash transfers most prominently but also subsidized health insurance, sometimes social pensions, and many more. PMTs are based on data analytics from household surveys but not on verification of household-specific information from existing government databases. The PMT method spread not just through a great deal of Latin America, but far beyond, sometimes to relatively similar settings (for example, the Philippines) and sometimes to far different ones, with many PMTs built in the lower income, lower inequality African countries.

PMT is something of a Rorschach test for those who think about targeting. Many, especially in ministries of finance or planning, see it as modern, scientific, data driven, replicable, and thus good for preventing patronage politics in social programs and safeguarding their reputation. Some communities find it a black box, a mystery. Analysts and observers have mixed opinions. Some find it a realistic, if imperfect, solution to a problem without perfect solutions; some find it anathema for its inbuilt statistical errors or lack of transparency. Some associate it with static survey sweeps and dislike those (although several countries using PMT have on-demand applications and dynamic registries). Many observers (the authors of this volume included) find that conversations about social policy proposals often jump far too quickly to issues around PMT, with insufficient discussion of the policy problem to be solved, the range of programming options, the range of possible eligibility determination methods that might be used, or how improvements in delivery systems could improve outcomes.

CBT is perhaps the oldest of the household-specific assessment techniques, but today, a much smaller share of programs apply CBT as a stand-alone approach. Conning and Kevane (2002) cite the use of CBT in historical events such as to support the 1834 English system of poor relief, in which local parishes performed some functions of local civil government, including the administration of poor relief, and the use of "native authorities" by the French and British as local leaders. In recent times, CBT is still among the most commonly used methods, especially in low-income countries, but it is rarely used alone and there are different ways of implementing it. In McCord's (2013) review of CBT experience (still the most recent comprehensive review), over half of the cases she identifies are in Africa, a plurality in Asia, a few in Latin America, and only one in Eastern Europe and Central Asia. She found information on 57 programs using CBT for which there were sufficient data for further analysis. In these programs, CBT was used alone in only three programs and paired with geographic targeting in half of the remainder. Some programs use CBT as a filter prior to using other methods, to narrow the pool of households still further. Others use community validation only at the end, taking steps to avoid reintroducing elite capture at this stage, such as allowing households

that were excluded during the original CBT process to be included or only allowing subtractions but not additions.

Despite the data revolution and increasing availability of other options, CBT remains a favored choice for its ties to the country political context. The community has been part of eligibility determination processes for years in many countries; thus, there is a strong sense that the community must still be part of the eligibility determination process. Beegle, Coudouel, and Monsalve Montiel (2018) highlight the importance of considering the possible trade-off between political and technical imperatives in designing targeting methods. In the limited comparative treatment arm experiments that have been done, CBT may be preferred by communities or lend cohesion.

CBT and PMT are sometimes used in the same program; a third of the CBT-based programs in ASPIRE also use PMT. There are multiple views on the logic of this, depending in part on the functions carried out by the community. Sometimes the community helps with functions such as outreach, prelisting households to be surveyed, and even data collection on a standard PMT form. The logic of using community members to complement and support administrative staff in conducting a PMT is clear, although in such cases, perhaps the nomenclature exaggerates the role of the community and it might be more accurate to label these instances as PMT methods. Where communities play a decision role jointly with a PMT, the logic is often stated as the community process guarding against errors of exclusion and the PMT guarding against errors of inclusion. This has a ring to it, but the logic is not unassailable. Running both processes may increase the costs and risk contradictions. In the Ghana study on this issue presented by Pop (2015), for example, CBT worked first to create a prelist of potential beneficiaries; then, PMT was brought to bear. By ruling out households put forth by the community, PMT lowered errors of inclusion somewhat, but it also introduced significant errors of exclusion with respect to the CBT-based lists. Moreover, in overruling the community, PMT undermines its power in decision making, and it may not abet the acceptability of decisions, hence creating some social tensions. Other evidence supports using CBT as a filter in cases where budgets or capacity are severely constrained or information sources limit the ability of more quantitative methods, such as PMT, to predict welfare status with enough accuracy. Adding PMT to a CBT process can reduce cross-community variation by bringing the following: (1) a common definition of poverty with more weight on money-metric poverty[8] to the process, (2) more training for community agents on the objective of the program; and (3) strengthened foundations of programs with preparation of clear operational manuals, information campaigns, and accountability mechanisms.

Another of the roots of current social assistance practice is humanitarian assistance, which is usually provided in response to some sort of emergency, financed by donors, and has a temporary vision and sometimes improvised methods. Because the programs are usually (initially) viewed as temporary,

building permanent administrative capacity may not be a priority. Indeed, often such programs are run outside government structures, with implementation reliant on international agencies, not-for-profits, and communities, although there is a movement to link humanitarian and development structures and methods more closely (see box 5.1). Humanitarian agencies

Humanitarian–Social Protection Alignment

Over recent years, a growing desire to move to more effective structures for addressing recurrent shocks has led to calls for greater coordination of humanitarian actors and government-led permanent social programs.[a] In Ethiopia, this led to the creation of the more stable and development-oriented Productive Safety Net Program and agreed systems to coordinate its expansion during droughts with humanitarian assistance. In Mali, the beneficiaries of humanitarian assistance linked to the political instability in the region of Gao and in the North of the country were eventually incorporated into to the national Jigisemejiri program so that they could benefit from the livelihoods component.[b] Hence, this coordination implies more stable and continuous support to the poor in "normal" times to build up their resilience and help mitigate the impacts of some of the recurrent shocks.

There is thus much discussion of how humanitarian and development assistance can work better together. Prominent threads of discussion include the use of geographic targeting and defining appropriate triggers for response to different shocks/crises and harmonization of eligibility criteria, in which the capacities of countries/agencies are built jointly or functions are shared between humanitarian partners and governments. Moreover, collaboration between governments and humanitarian agencies in implementing multiyear government programs can generate the resilience needed by the population, as in some cases, the government may have limited capacity to operate in certain areas of the country, particularly in the case of conflict. So far, joint systems are more the exception than the rule, but such collaborations can be seen in a few countries as presented by Gentilini, Laughton, and O'Brien (2018), who review 12 country case studies.

The humanitarian context raises several context-specific design issues. For example, have livelihoods been lost to a disaster and is immediate cash or in-kind support needed? Will access to documentation and work permits enable displaced people to access employment?

BOX 5.1 *(continued)*

If income support is provided, how do the benefit levels compare with average household consumption of the poor? This can matter particularly in displacement contexts. When internally displaced people receive significant humanitarian support while other conflict-affected but nondisplaced households do not, social tensions can arise, as in Iraq.[c] Similarly, relatively generous support for Syrian refugees in Iraq, Jordan, and Lebanon has created tension when poor households in host communities receive much less support from government programs (World Vision 2015; Durable Solutions Platform 2019).

A comparison of targeting approaches can be the start of understanding how to align humanitarian and government assistance systems. In Iraq, a desk study compared the government's social assistance proxy means testing with that of allied humanitarian organizations. This was particularly relevant for two reasons. First, government social assistance had not been operating in territories controlled by the Islamic State of Iraq and Syria (ISIS), which were an important focus of humanitarian activity. Thus, the humanitarian database was potentially a valuable source of referrals to government programs during recovery. Second, the overlap in eligibility for the two systems could inform the budgetary needs of expanding government programs. The desk study found a strong degree of overlap, which is currently being confirmed through collection of new data in the field.

a. In low-income/high-poverty countries that are highly dependent on rainfed agriculture, droughts are recurrent facts of life and the population has little resilience, so cycles of temporary emergency programming are recurrent.
b. Premand and Stoeffler (2020) show that a multiyear government cash transfer program in rural Niger increased household consumption by about 10 percent on average among households affected by recurrent drought shocks. The transfers increased savings and helped households protect their earnings in agriculture and off-farm businesses when shocks occur.
c. https://www.brookings.edu/blog/markaz/2015/04/28/the-overlooked -humanitarian-crisis-in-iraq-the-need-to-address-disparities/.

developed the Household Economy Analysis (HEA)[9] approach, a livelihoods-based framework, and used it to anticipate and respond to food crises. The HEA baseline defines livelihood zones, that is, geographical areas within which people broadly share the same patterns of access to food, income, and markets (Holzmann et al. 2008). The approach is sometimes used to identify geographic zones that are at risk of food insecurity. Within geographical areas, it identifies different wealth categories with clearly defined household characteristics, assets, income, and food needs. This is sometimes used for household-specific assessments as well, although the practical way of doing this can vary. In some places, a pure CBT process is used to identify

households in each wealth category and target beneficiaries for assistance. In other places, focus groups are used to identify the characteristics of households in each wealth group and inform weights that can be used in a formula-based approach (Schnitzer 2019). Utilization of HEA is prominent among humanitarian agencies in the Sahel, but it is not widespread globally.

Public lotteries are sometimes used to allocate rationed slots in a program. Bance and Schnitzer (2021) challenge the social protection community to consider when lotteries may be useful. The view of the authors of this book is that a public lottery is a complementary tool for supporting the final selection of beneficiaries when there are more eligible people than can be covered; it is not a targeting method in and of itself since it does not measure or rank any aspect of welfare. Public lotteries have been used especially for public works programs (cash or food for work) in places as diverse as Argentina, the Central African Republic, the Democratic Republic of Congo, Sierra Leone, and Tanzania, which, as the list suggests, are mostly low-income and/or fragile settings.[10] This type of program is perhaps used more for public works than straight cash transfers because with cash transfers, it is easier to avoid rationing by spreading low benefits among many people. For public works, there is something of a lower bound set by wage levels and the nonwage costs of the programs, so the need to ration is harder to avoid. The process of running a lottery is quite simple. Once an initial list of applicants is defined using any of the aforementioned targeting methods, a random selection process is used. Program administrators can run the selection using electronic random draws or by picking numbers out of an urn at a community meeting. The principle of the fairness of the process is considered a plus, especially in fragile and conflict settings, as applicants have an equal chance to be selected. Errors of exclusion are clearly present as all those in the lottery are deemed needy/eligible according to the self- and geographic-targeting criteria, but the errors are transparent to political processes.

With so many different influences and contexts, it is not surprising that the choices of targeting methods, their implementation, and outcomes are varied. Although this section has described patterns that exist, there is enormous variation in implementation. It seems that every region has a few cases that do not fit the pattern, and virtually every combination of methods has been used somewhere, sometimes in unlikely places.

A common thread may be increasing expectations. When systems of eligibility assessment started to be developed, the role of social protection in development was much less accepted and capacities and expectations were rather low. The gradual building of capacity in countries that once had none has inspired those that more recently began the journey, and the data revolution inspires all. With this, the emphasis has shifted. If in prior

decades it seemed that it was progress to provide direct support to some of the poor who before that had none, now it seems insufficient not to have helped them all.

Considerations in Choosing among Welfare Targeting Methods

Choosing a targeting method requires matching different methods to a program's objectives, particularly whether beneficiaries will come from an entire category of households or be selected from a welfare ranking of households:

1. *Are household-specific ranking and measurement needed or will another method or methods suit the purpose and context?* Will one or a combination of self-targeting, geographic targeting, demographic targeting, or even lotteries fit the purpose and constraints well (or well enough)?
2. *If a household-specific method is needed, which should be used?* For example, which among means testing, HMT, PMT, or CBT is most suitable?
3. *Should multiple methods be used?* It is now quite common that multiple methods are used, but it is not necessary and sometimes not ideal. For example, although methods that determine eligibility based on a category can help prioritize when budgets are limited or sequencing is required, they are guaranteed to exclude some poor people, as poverty affects people in all places, all ages, all genders, or other group definitions. So, if a country is developing or has developed the capacity for a household-specific system, does a categorical system add value?

Practical considerations also influence the choice of methods. When choosing a targeting method, policy makers will also want to consider what data are already available and what capacity public agencies have for using them, especially when household matching is required across databases without unique identifications (IDs). Capacity concerns may also arise for more technical targeting methods, including the development of poverty maps for geographic targeting and household income imputation techniques underpinning HMT and PMT. In addition, the available budget for beneficiary outreach and selection may make some methods less attractive, especially if large-scale data collection is required through home visits.

One of the powerful issues is whether a program is for "good times" or shocks/crises or both. As shocks are quite diverse and have different impacts on different populations, there is no unique targeting method that could be used against all shocks equally well. Speed of response is of the essence for shocks. Programs that address chronic poverty or redistribution can build their capacity or more easily take time to gather household-specific data.

Chapters 3 and 4 cover various considerations for differentiating eligibility around covariate shocks but mostly with respect to planning and delivery systems and thus in a manner relatively neutral to methods. As this chapter examines each method, it adds a few notes that are pertinent to this choice, including each method's relevance to idiosyncratic shocks to individual people or households.

Finally, there can be a degree of path dependency, although history is not destiny. It is perhaps easiest to build on existing administrative capacities or not to disturb existing consensus or institutional arrangements. In some cases, such factors can lead to a different choice of methods than might occur by starting from scratch. Over time, countries can move from using only simple methods to more complex ones as they build their social protection programming and delivery systems (see box 5.2).

BOX 5.2

Djibouti: The First Steps toward a Targeting System in a Fragile State

At the time of the response to the food, fuel, and financial crisis in 2008, Djibouti had neither a flagship social assistance program nor a social registry. The principal instrument used in response to the food crisis was the elimination of taxes on selected food items, a rather blunt instrument. Over the years, various programs were established in the wake of drought shocks; these were largely time-limited, donor-driven initiatives, working outside government systems, and mainly focused on providing food to vulnerable populations.

As Djibouti began to expand its social assistance programming, in 2010, it launched the Employment and Human Capital Safety Net (EHCSN) project, implemented by the Djibouti Social Development Agency with support from community-based organizations, nongovernmental organizations, and small and medium-size enterprises. The EHCSN was a workfare program with a nutrition component focused on nutrition education and provision of micronutrients. The Employment and Human Capital Safety Net project applied multiple simple targeting mechanisms. It used geographic targeting based on poverty maps. Participation in public works implied self-targeting. Demographic criteria focused the program on "nutritionally vulnerable households"—those with a pregnant or lactating woman or

BOX 5.2 *(continued)*

a child younger than age two in the household. Communities helped to identify which households met these criteria. An initial assessment indicated that 73 percent of the beneficiaries were poor.

Another milestone in the development of social assistance came in 2016 when Djibouti introduced the National Programme of Family Solidarity (Programme National de Solidarité Famille [PNSF]), a conditional cash transfer. For this program, community-based targeting was used in rural areas. A proxy means test was used in urban areas as it was deemed that community cohesion was not as high in urban areas (communities remained involved in the delivery system and mobilization activities).

In parallel with this new program, the government developed the Djibouti Social Registry, managed by the Ministry of Social Affairs and Solidarity, which now covers about half of the population. After initial en masse registrations, new households can register themselves in one of 12 Ministry of Social Affairs and Solidarity branches across the country. The social registry includes the capacity for including biometric information on individuals benefiting from the programs, which is used to verify that no duplicate benefits are provided by the programs using the registry. Currently, 13 programs share the social registry data for eligibility assessment, most prominently for the universal health care program, the provision of social housing, the provision of microfinance, and PNSF. The Djibouti registry was established by a law (decree 2017-311/PR/SEAS) in September 2017. In 2020, the government began registering refugee households in the social registry, permitting them to be considered for eligibility for the programs that use the registry for eligibility determination.

The value of the social registry became apparent as it permitted a rapid response to the COVID-19 pandemic, facilitating targeting and enabling the rapid deployment of in-kind transfers during the pandemic lockdown. Just weeks after the onset of the pandemic, the government put in place a program to provide food vouchers to 90,000 poor and vulnerable urban households. Using the social registry, the intervention targeted households under the poverty line and those active in day labor or temporary and/or independent work—reaching about half of the country's population. Vouchers worth DF 30,000 (US$170) entitled beneficiaries to a basket of food staples. At the same time, the PNSF coverage was expanded, bringing the total program coverage to about 9 percent of the population.

Despite fragility and limited capacity, Djibouti developed a long-term strategy to create the foundations of a permanent social

BOX 5.2 *(continued)*

assistance system that would strengthen the country's ability to respond to future shocks as well as build human capital and fight chronic poverty. The initial program was designed to address emergency needs, but it evolved based on assessments, with coverage increasing as programs matured. Djibouti has shifted from only self-targeting (for food subsidies) to both geographic and demographic targeting (for public works) to household assessments (for cash transfers and food vouchers).

Sources: Mendiratta et al. (2020); World Bank (2013). See also Brodmann, Devoto, and Galasso (2015); Devoto, Galasso, and Brodmann (2017); Leite et al. (2017); Machado et al. (2018); UNICEF (2010).

The rest of this chapter considers program objectives and practical or contextual factors when choosing targeting methods, beginning with the former. Figure 5.2 shows the decision process for selecting one or more targeting methods. The first question is whether a program's objectives determine household eligibility by self-targeting, category, or an attempt to rank households on a money-metric welfare basis. Methods for the former are discussed first, including self-targeting, geographic targeting, and categorical targeting. The options for ranking household eligibility by welfare are discussed next, covering means tests, HMTs, PMTs, and CBT. These two choices—categories or rankings—are not mutually exclusive; locations can be selected by geographic targeting and households within those locations ranked to select beneficiaries. Mixed methods are the most common approach in practice. As each method is covered, the second set of practical considerations are also discussed. The text also pays attention to the hard cases: places with low inequality where it is difficult to distinguish between households, places with low capacity to implement different targeting methods and low budgets to build capacity, and places with conflict and displacement that may be particularly sensitive to social discord. In many cases, these difficulties coexist in the same place.

Self-Targeting Methods

Self-targeting programs are open to all, but they are designed in such a way that they are used disproportionately by the poor. The nonpoor choose of their own accord not to use them. The factors that contribute to this choice are preferences about quality, private or transaction costs of participation, and possibly stigma associated with the use of the service or program. The basics have been well known for years (see, for example, the treatment

Figure 5.2   Factors to Consider in Choosing a Targeting Method
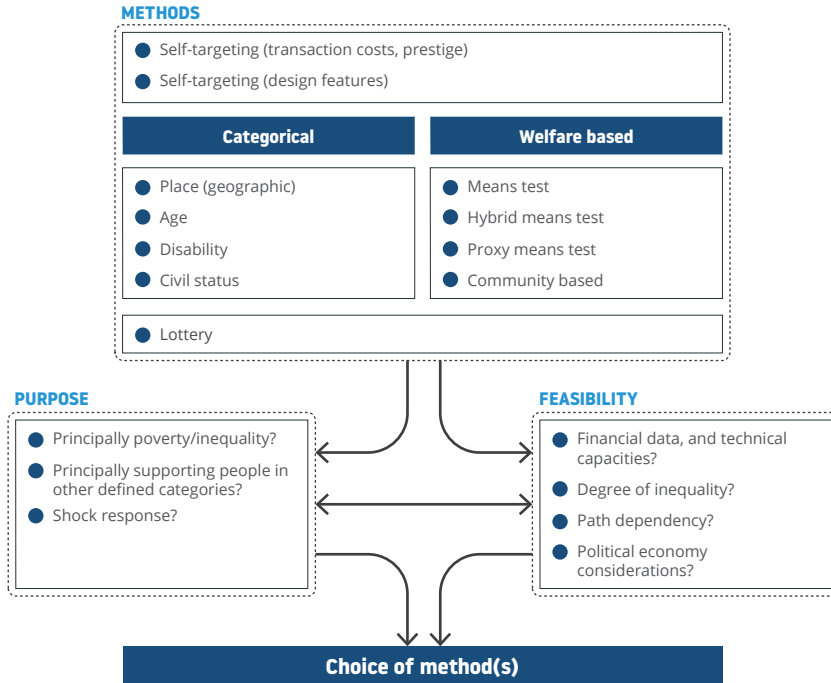
Source: Original compilation for this publication.

within the targeting literature in Besley and Coates [1992]; Besley and Kanbur [1988]; Coady, Grosh, and Hoddinott [2004]; and Devereux et al. [2017] or the social assistance literature, including Grosh et al. [2008] and Pinstrup-Anderson [1988]) and considerations about self-targeting have changed less than for other methods. The chapter treats the method some-what lightly for that reason but provides a briefing to round out the treat-ment of choice of methods.

A classic and still prevalent application of self-targeting is the offer of work on public works jobs for a low wage. According to *The State of Social Safety Nets*, nearly 100 countries operate such programs (World Bank 2018b). Figure 2.6 in chapter 2 shows that the incidence of public works programs is in the same range as other types of programs, although few rely only on self-targeting. Among the attractions of self-targeting through a work requirement is that the work requirement may increase the political support for the program, overcoming the notion that nonworking adults are not deserving.

To ensure that the program favors participation by the poor, the work is low skilled and low paid. The work in such temporary employment programs is also commonly, although not necessarily, physically strenuous and performed outdoors in the creation or maintenance of small-scale infrastructure.[11] Rwanda has an expanded public works program that includes paying women to provide childcare in their communities (World Bank 2020d). Gaza's public works program focuses on providing social services (World Bank 2018c). The wage is a key variable; regularity of work and working conditions also matter. Obviously, the lower the wage is set, the lower is the chance that nonpoor persons would sign up. In the conceptual ideal, the program may pay just a little less than the market wage for similar work as a way to balance the goals of self-selection and adequate pay. There are various brakes on setting a very low wage. First, if it is set too low, the benefit to the worker would be little, especially if they have to lose some hours or days of work they would otherwise do. Estimates of such forgone earnings vary widely depending on the setting, ranging from 7 to 50 percent (Subbarao et al. 2013). In general, countries try to time works to seasonal down periods, for example, the slack agricultural season, and sometimes have relatively short workdays to minimize forgone earnings. There may also be legislative barriers to paying less than the minimum wage for the country, although if market wages surpass minimum wages as they do in many countries, this will not be a concern. In Subbarao et al.'s (2013) survey of public works programs, the majority pay less than the market wage.

Self-targeting via low-wage work may be sufficient to ensure that applicants are relatively poor, but quite often the size of the program is too small to accommodate all the applicants and so ancillary mechanisms may be used. Public works schemes typically cap the duration of public works employment to share the benefits more widely and avoid attracting laborers with steady employment. It is also common to see elements of geographic targeting of the public works. Sometimes other methods are used—categorical (as in the case of Djibouti in box 5.2), CBT (for example, in Ethiopia and Rwanda), lottery (for example, in the Democratic Republic of Congo), or occasionally PMT (for example, in Tanzania).

As self-targeting via low wages for public works may exclude some needy households or individuals, care is needed to minimize this. Some of the poorest people typically live in households with few or no working-age adults or adults whose work is limited by caregiving responsibilities, social norms, or disability, although there are some approaches to lower such barriers. Burkina Faso provides childcare for female workers on traditional public works.[12] India's Mahatma Gandhi National Rural Employment Guarantee Act program requires that work be located within 5 kilometers of each claimant's home, which bounds commute time and the cultural challenge of women being far from home; moreover, the program pays

equal wages to women and men, despite prevalent wage gaps in the private sector. Various countries have minimum quotas for women's participation (see, for example, Curry 2019). Malawi has been piloting ways to improve the access of persons with disabilities to jobs on public works projects (Vikan and Diekmann 2017)**.** South Africa sets aside 2 percent of assignments for persons with disabilities (Letswalo 2020).

Another classic self-targeting method is subsidization of the prices of basic food stuffs, ideally of foods that are more consumed by the poor than the nonpoor. The idea is to find different staples or variations on them that are nutritionally equivalent or closely so but differ in terms of prestige—sorghum versus corn, broken rice versus whole, coarse flour versus fine, and yellow versus white corn are examples in which the former is usually less prestigious but (at least) as nutritionally equivalent as the latter. If the price of the less desired commodity is subsidized enough, the poor who are still trying to meet their caloric needs will buy it, while the nonpoor will purchase the more prestigious variant. Of course, the sorting will be inexact and dependent on the relative strengths of people's preferences and the differences in prices. Moreover, there may not be a commodity that is consumed more by the poor than the nonpoor (especially if this is judged in absolute terms rather than in relative ones). Even if there is one such commodity, it needs to have a production and trade chain that makes it easy to attach the subsidy. For example, grain grown by smallholders and sold in a thriving private market to dispersed outlets will be harder to subsidize than a product that is largely imported by a monopoly state trading agency. Consumption patterns are important as well. Sorghum or millet, for example, may be consumed not only by poor humans, but also used as animal feed. Thus, subsidies on these grains may result in a costly indirect subsidy to the livestock industry. There may also be regional variations. The urban poor may purchase tortillas daily, but the rural poor may make them at home.

Whether the benefits of food subsidies are close to neutral in their distribution or somewhat regressive depends on the commodity that is subsidized and patterns of consumption. The logic of why subsidies on food staples can be reasonably self-targeting is intuitive—even an overfed nonpoor person will only eat a modest amount more rice, bread, or porridge than a poor and hungry person. The richer man will diversify his food basket to more nutrient dense, tasty, convenient, or luxury foods and his consumption basket to a smaller share of food with more nonfood, while a poorer person's food basket will remain more concentrated in the staple grain. Thus, a subsidy for rice may result in rather flat incidence. A subsidy for sugar or meat would have a more regressive incidence because the rich person can more feasibly eat more sugar or meat than the scant amount of these that the poor eat. In the Arab Republic of Egypt, for example, in the 2000s, about 96 percent of the poor were benefiting from the food subsidy

system, but those in the richest quintile were receiving about 12.6 percent more from food subsidies than those in the poorest quintile. The *baladi* bread and cooking oil subsidies were the most regressive of all the food subsidies. The only subsidized food that was progressive was *baladi* wheat flour, which provided the poorest quintile as much as six times the benefits as the richest (World Bank 2010a). Because food price subsidies are a relatively blunt tool, some programs also use some other rationing or eligibility determination tool. For example, Indonesia's variously named rice price subsidy first used CBT and then PMT (Holmemo et al. 2020), and India allows the poor larger purchases at discount prices in its Public Distribution System (Dreze et al. 2018).

As practiced, most food subsidy programs leave the government defending a set price with high upward risk for fiscal costs and political risks for price changes. Sometimes there are significant leakages into the black market. Thus, there is a long history and rich literature on attempts to reform food price subsidies (Pinstrup-Anderson [1988] is an authoritative source). For example, Tunisia engaged in a multiyear reform of food subsidies, which saved 2 percent of gross domestic product (GDP) by reducing the range of foods subsidized and shifting the degree of subsidy among nutritionally similar items (for example, eliminating subsidies on white flour *baguette* but maintaining them on coarser flour *gros pain*; liberalizing the market for fine olive oil but subsidizing generic grain-based cooking oils sold in small qualities with bring-your-own bottle packaging, and so forth (Tuck and Lindert 1996). Alderman, Gentilini, and Yemtsov (2018) provide multidecade treatments of the stories of reforms in Egypt, India, Indonesia, Mexico, Sri Lanka, and the United States. Eventually, Mexico moved to cash transfers in lieu of the food subsidy programs; Indonesia reformed its program many times and, in a subsequent move, converted to a food voucher (Holmemo et al. 2020); the United States maintained the food stamp system, eventually using quite sophisticated targeting and payment mechanisms; and Egypt and India reformed, reduced market distortions, and improved governance of their subsidized rations systems but still maintain programs with very high rates of coverage and significant budget.

Energy price subsidies are still common and tend to be much more markedly regressive than food subsidies. The intuition is again simple—everyone consumes food and there are limits to how much more staple food a wealthy person can eat than a poor one, but there is no such analogy for fossil fuels. The poor may consume no or very little fossil fuel, and the rich may consume a great deal because they have many appliances, possibly air conditioning or an automobile. In the Islamic Republic of Iran, prior to subsidy reform, the government spent about 20 percent of GDP on food and energy subsides. The prices of bread and flour were subsidized, with the value of the subsidies essentially flat across the income distribution, whereas the value of energy subsidies was more than five times higher in

Figure 5.3 Percentage of Energy Subsidies, by Household Decile, Selected Countries
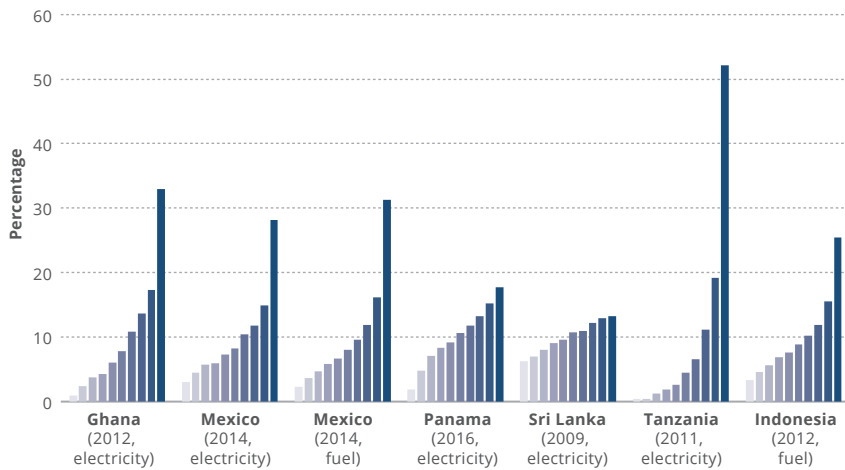
Sources: Based on CEQ Institute database.

the highest decile than the lowest. Silva, Levin, and Morgandi (2012) provide other examples of flattish food subsidy incidence and markedly regressive energy subsidy incidence in the Middle East and North Africa. Figure 5.3 shows highly regressive electricity and fuel subsidies for selected countries in the year of analysis (several of the countries have since implemented various reforms and changes in pricing regimes; for example, Indonesia has transitioned to an expanded targeted direct transfer system).[13]

There have been some attempts to subsidize some fuels more than others, to favor those more used by the poor, yet with little success. For example, kerosene is often subsidized to help the poor use it rather than biomass for lighting or cooking. But this commonly results in commercial malpractice, such as kerosene being used to dilute diesel or diverted to the aviation sector, or various black market, smuggling, or "commercial tourism" schemes (see Kojima 2013).

Moreover, energy subsidies can promote overconsumption of energy, usually that derived from fossil fuels, and are thus inefficient and harmful to the environment and a target for policy reform for multiple reasons (see, for example, Coady et al. 2015; Flochel and Gooptu 2017). Kojima (2013) documents how 65 countries struggled with energy subsidies in the years around and following the 2009 fuel price crisis. Kojima (2021) provides a recent overview and nine case studies around the reform of liquefied petroleum gas (LPG) subsidies, including several well-known cases that moved to household-specific alternative targeting mechanisms for compensatory cash transfers.

Overall, the price subsidy experience underscores the value of being able to go directly to households, although often with fairly high thresholds. India's LPG reforms and the Islamic Republic of Iran's comprehensive reforms started with universal cash compensation (Gelb and Mukherjee 2019; Salehi-Isfahani and Mostafavi-Dehzooei 2018). Jordan's fuel subsidy reform gave compensation to two-thirds of households (Atamanov, Jellema, and Serajuddin 2015), as did El Salvador's LPG reform, the Dominican Republic's somewhat lower at about a third of the population, and Brazil's closer to a fifth (Kojima 2021).

Categorical Methods

Several methods work by assigning people to a category based on easy-to-observe characteristics and assuming that needs are relatively homogeneous among the group. The most commonly used methods are geographic targeting and demographic targeting. Geographic targeting is quite simple to understand conceptually, although the data and inference issues are more complex. Thus, it gets short treatment in this chapter; more details are provided in chapter 6.

Geographic Targeting

Some variant of geographic targeting is applicable in many settings, although two of the options are designed to cope with rationing of programs in ways that it is hoped will diminish as social protection systems continue to develop.

- First, the most extreme variant of geographic targeting selects areas where the program will operate and gives benefits to all in those areas. This may be highly pertinent in situations such as natural disasters that affect only some areas with widespread losses.
- Second, in a more common variant, the neediest areas are selected as places where programs will operate, with the decision based on some spatial analysis indicator(s) related to need—such as poverty, drought, or malnutrition that is pertinent to the purpose of the program—and then additional eligibility criteria are used within the areas of operation to select the households that will benefit. The exclusion of whole parts of the country leads to clear and politically visible errors of exclusion since poverty and most other vulnerabilities occur everywhere, just at different rates (see box 5.3 for an example). This variant is often chosen where need is highly geographically differentiated and/or for programs that are in a roll-out phase.
- In the third variant, the program operates in all geographic areas, but geospatial analysis is used to define benefit quotas for each area, with specific households selected via some other method. When used to allocate the caseload or places in a budget rationed program, there will still

BOX 5.3

Ethiopia's Rural Productive Safety Net Program: Geographic Targeting

Ethiopia's rural Productive Safety Net Program (PSNP) is mostly a public works program, with direct support to labor-constrained households that cannot easily fulfill the work requirement. The labor is used on works that provide local public infrastructure and services, with a heavy emphasis on watershed rehabilitation. The PSNP was launched in 2005 just after a major food crisis and at a time when there was significant concern about the ability of emergency food assistance to address growing chronic food insecurity in rural areas.

Given this evolution, the program is geographically targeted: only districts (known locally as *woredas*) that received food aid for the three consecutive years preceding the launch of the PSNP were included in the program's areas of action, about 40 percent of the woredas in the country. Within woredas, there was further geographic targeting to select the subdistricts (*kebeles*) that would participate. This was the responsibility of the woredas and historical receipt of food aid in the kebeles was one of the common criteria used. This metric corresponded to the policy concern and fit within the constraints of the data available at fine levels of disaggregation at the time the program was launched.

The geographic targeting of the PSNP illustrates some of the perennial considerations about geographic targeting. At the time of its launch and throughout, the PSNP's budgets and caseloads have been rationed in stringent ways. Moreover, the program is complex to implement. Thus, focusing the caseload and development of institutional capacities in a very constrained environment may be a sensible, if difficult, triage decision. The geographic focus of the program on food insecure woredas also established a politically acceptable means of rationing resources in a country with a large poor population; conversely, it also provided a basis for sequencing the addition of woredas into the program, with woredas meeting the criterion of receiving food aid for the three consecutive years being added to the PSNP in 2006 and 2010. The selected woredas are less food secure and poorer, and their households have fewer durable goods and livestock than those in non-PSNP woredas. Most marked is how much worse the vegetative index was than in non-PSNP woredas (see figure B5.3.1). However, because the program inherently excludes many districts and has a budget-rationed caseload, there are significant errors of exclusion. In 2016 (the most recent household survey analyzed in the 2020 poverty assessment), only 13 percent of the poor population in Ethiopia was covered by the PSNP.

BOX 5.3 *(continued)*

Figure B5.3.1   Greenness of Woreda Vegetation Is the Strongest Correlate of Woreda Selection for the PSNP

Standardized mean differences between PSNP and non-PSNP woredas, 2016
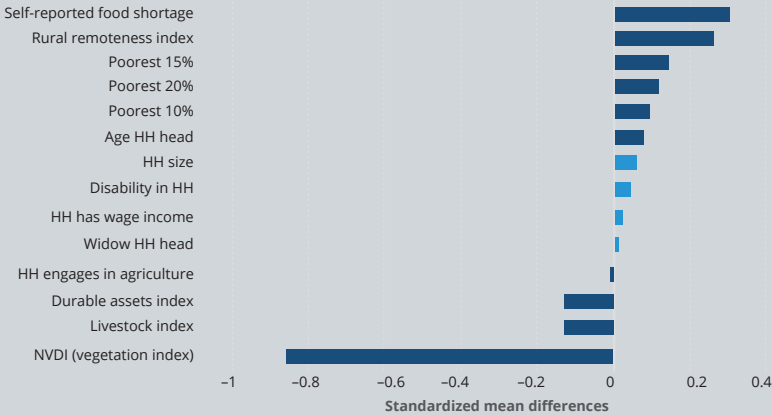
Source: Calculations using data from the Household Consumption and Expenditure Survey (HCES) and the Welfare Monitoring Survey (WMS) 2016.

Note: The bars in dark blue are statistically significant at the 10 percent level or lower. HH = household; NDVI = Normalized Difference Vegetation Index; PSNP = Productive Safety Net Program.

There have been changes in economic context over the life of the program and in the government's policy commitments. Although it is still considered a low-income country, Ethiopia has had high growth rates in recent years. Extreme poverty (as measured by international poverty lines) fell from 56 percent in 2000 to 24 percent in 2016. The prevalence and depth of food insecurity improved even more. In 2005, close to one-third of Ethiopians reported experiencing food shortages in the 12 months preceding the survey. This decreased to 22 percent in 2011 and 10 percent in 2016. Overall, the number of months per year an Ethiopian citizen experienced food insecurity decreased from 1.2 months in 2005 to 0.7 month in 2011 and 0.3 month in 2016. From the impact evaluations, it is clear that the PSNP has substantially contributed to these improvements.

The improvements in food security overall began to shift the focus of policy dialogue to include not only food security, but also a more general definition of poverty as the basis for establishing program eligibility—an evolution reflected in the national social protection policy, which was adopted in 2014. By this metric, the selection of woredas

BOX 5.3 *(continued)*

contributes little to the PSNP's targeting performance as poverty is widespread rather than confined to a subset of woredas (figure B5.3.2). The second-level selection of subdistricts (kebeles) within the participating woredas contributes more to poverty targeting, and the selection of households by community-based targeting contributes still more.

To reduce the errors of exclusion in the geographic targeting of the PSNP, the government is working on two fronts. First, in 2016, the Urban PSNP was launched. To update the geographic targeting in rural areas and reduce structural errors of exclusion, in 2020, the government decided to revise the allocation of caseloads among woredas, including additional woredas in the program and reallocating the caseload among participating woredas. The new-to-the-program woredas will use a combination of (1) recent history of receipt of drought-related emergency food assistance, (2) remote-sensing satellite data showing the frequency of drought shocks, and (3) the prevalence of extreme poverty. The redistribution of the existing caseload among new-to-the-program and already participating woredas will be informed by poverty data, but it may take other factors into account, including high vulnerability rates in certain regions and/or the political risks of

Figure B5.3.2 Woreda Selection Does Not Add Much to the PSNP's Targeting Performance

Decomposition of the targeting differential
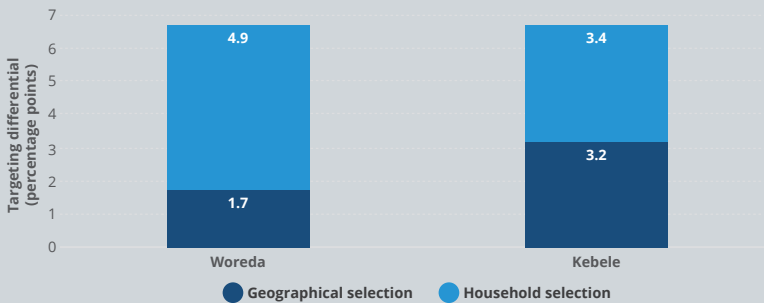
Source: Calculations using data from the Household Consumption and Expenditure Survey (HCES) and the Welfare Monitoring Survey (WMS) 2016.

Note: The targeting differential is the difference between coverage of the poor and that of the nonpoor. PSNP = Productive Safety Net Program.

continued next page

BOX 5.3 *(continued)*

significant caseload reductions. At a minimum, this redistribution will achieve the following: no regional caseload should exceed the projected number of people living below the 15 percent poverty line by more than 10 percent; regional caseloads will be primarily redistributed within the respective region (to new-to-the-program and previously covered woredas) according to the poverty data; and expansion will be planned for a minimum of 70 new woredas.

In the future, such data-driven reallocation of the caseload is planned to take place every four years. Thus, the PSNP continues to use geographic targeting but with the criteria, process, and locations refreshed to match changes in context and the availability of data.

Sources: World Bank 2015a, World Bank 2015b, World Bank 2020c, World Bank 2020e, World Bank 2021a.

be errors of exclusion due to budget, although they will be more subtle and distributed across subnational polities.

- In a fourth variant, there is no rationing, but geospatial analysis is used in program planning and monitoring to ensure good outreach, optimal allocation of staff, and so forth to reduce transaction costs for clients. The program per se is not geographically targeted, but administrative resources are geographically targeted in an effort to minimize errors of exclusion.

In each variant of geographic targeting, it is necessary to consider the technical and political process by which the budgets will be allocated and updated over time as patterns of need change, new data are collected, program coverage or impacts evolve, and so forth. As chapter 6 discusses, whether traditional survey-to-survey imputation methods are used or newer approaches using big data are employed, the technical requirements to guide geographic targeting are significant. It is now possible in most cases to provide reasonably up-to-date poverty estimates at low levels of disaggregation, usually the third level of administrative unit (for example, the county, parish, or subdistrict) and sometimes even more finely (village or urban neighborhood). The political factors may be significant as well. If the poorest or most vulnerable areas are concentrated in a few of the second-level subnational jurisdictions (for example, states or departments used in defining representation in the national legislature), then there may be tension between allocating the program or program slots to the poorest or most vulnerable third-level units observed in the geospatial analysis over the whole nation (for example, the poorest 100 districts), or allocating the

program or program slots so that all or most legislative districts have some and the slots are allocated to the poorest areas within each legislative district (for example, the poorest 10 districts in each of 10 departments). Reallocations over time will again have a political element as some areas/jurisdictions would gain slots and others lose as the geospatial distribution of poverty or vulnerability changes over time.

Geographic targeting will clearly be suitable as part of response planning for a large share of natural disasters as these are normally geographically delimited (although for small island states they may encompass the whole country). In contrast, geographic targeting is not particularly useful for widespread economic shocks nor for individual health or employment shocks.[14]

Demographic Targeting

Demographic targeting is used in half or more of social assistance programs. (As coded in the ASPIRE database, categorical targeting is used in three-quarters of the programs, and about three-quarters of these cases are probably using a demographic category.) Around a third of the categorically targeted programs (not all of which are demographic) also assess eligibility with another method. Looking at cases of categorical programs with two or more methods, the most common additional methods are geographic (43 percent), means testing (40 percent), and PMT (25 percent).

Demographic targeting demands careful thought about the unit of assistance. Society has a dual concern for both individuals and families. Discourse about human, social, and economic rights is framed around individuals. Yet, since time immemorial, families have been the way in which resources and risks were shared and most people live in families, so it is hard to divorce the discussion of individuals from the context in which most live.

The following list unpacks the appeal of demographic targeting. Although in some cases it is difficult to quantify, several of the advantages of demographic targeting are generally accepted (Devereux et al. 2017; HelpAge 2006; UNICEF-ODI 2020) and universal child allowances and social pensions are included in the International Labour Organization's vision and costing of social floors (Durán-Valverde et al. 2020; ILO 2019; Ortiz, Cummins, and Karunanethy 2017). The following are among the advantages that are commonly referenced:

- *Political consensus for supporting the meritorious or deserving.* The groups supported through demographic targeting are generally viewed as deserving–children are to be treasured and protected, and their human capital and future are highly sensitive to any deprivations. The elderly are to be respected and rewarded for their life service. Widows have suffered misfortune in losing their life partners.

- *Stigma, transparency, and human rights.* There is a consensus that age-based targeting carries no stigma. Age is a natural state of life, not subject to any lack of initiative by the individual (or even family). Indeed, families are congratulated on the good fortune of a new birth and the elderly on the good fortune of a new birthday. Demographic targeting also has the positive feature of transparency, making it easy for claimants to understand what is due and seek redress if they have not received it. Demographic targeting and its ability to meet the principles of equality and nondiscrimination are among the most acceptable under human rights critiques.
- *Demographic targeting is simpler to implement than other forms of targeting.* Programs that use only demographic targeting have lower information requirements than programs that try to measure or estimate welfare, employability, or disability to establish eligibility. However, programs that are only demographically targeted still need the whole delivery chain elaborated in chapter 4—outreach, intake and registration, payment, recertification, grievance redress, and monitoring. These functions must operate continuously so that each new cohort of births and birthdays is accommodated and with low transaction costs. If these functions are not done well, there may be errors, especially of exclusion, and sometimes of inclusion and/or loss of reputation for the program.[15]
- *Less direct concerns about labor disincentives.* Sometimes the political consensus to provide social assistance to those in need is a bit frail, and/or concern over labor disincentives is strong so that programs for those not expected to be able to work—children, the elderly, and those living with disability—achieve more significant funding. (Although this phenomenon seems strong, it seems to undervalue the evidence that labor disincentives are usually minor and gives little acknowledgment that most individuals in the favored groups live in families that share resources and most of those families contain working-age adults whose employment and time use decisions may be sensitive to unearned income.)
- *Empowerment within the family.* A strand of the literature clearly recognizes the family and that while there may be some pooling of resources, not all individuals have equal voice or share. This strand of the literature reminds that having an independent income stream can elevate the status of the elderly within the family (see, for example, Kidd 2016; Tran, Kidd, and Dean 2019).

Sometimes demographic targeting is a good fit for purpose for a program. Vaccination of young children is a good example. Individual children benefit from gaining immunity as early as they can. There is a benefit to others in the community as each child's vaccine helps lower the potential for the illnesses to spread in the community.

Sometimes demographic targeting is used not as perfectly fit for purpose for an inherently individual- and age-related service or benefit but as a pragmatic way of serving poverty reduction or risk management functions. For poverty, the line of reasoning is that if children and the elderly do not work, then their supporting families will have higher dependency ratios than others and thus be poorer on average. For risk management, with respect to the risk of outliving one's earnings or savings, governments the world over mandate contributory pensions. But their coverage is often insufficient to provide old-age income security broadly, and they are sometimes highly subsidized. Thus, if it is desired to make subsidies for old-age support transparent or ensure coverage of those who do not contribute via labor/social security taxes but rather to general revenues, social pensions may be an important strand of pension policy (see, for example, Packard et al. 2019).

It is therefore of interest to understand the correlation of different demographic categories with poverty. The correlation varies by group and from place to place but in general is mild. This finding is not new (Guven and Leite 2016; Kakwani and Subbarao 2005; Slater and Farrington 2009), but it is worth looking at updated data and examining the patterns and variability within them. For this, household survey data from the World Bank's Global Monitoring Database are used.[16] The potential results for using the demographic category to proxy poverty can be compared with other methods in a specific country in simulations, a topic taken up later in the chapter.

The number of people in a particular category varies considerably across countries, especially by income level; this is an important factor that will likely drive both the possible coverage of the poor with demographic targeting and the budget requirements. In poorer countries, children make up a large fraction of the population, on average nearly 50 percent of the population in low-income countries and 35 percent in lower-middle-income countries (table 5.2). This means that programs targeting children are likely to cover a significant share of poor households, but to have large overall coverage thus requires commensurately large resources. The elderly are

Table 5.2   Share of the Population, by Demographic Category (%)

Income category	Children under 5 years	Children 6–14 years	Adults 15–64 years	Elderly 65+ years	Widows under 65 years
Lower	20	26	51	3	8.6
Lower middle	14	21	57	8	8.1
Upper middle	7	11	72	10	4.6
High	6	10	65	20	0.2

Source: Calculations using data from the Global Monitoring Database.

much less numerous, as only 3 percent of low-income country populations are elderly on average. Thus, programs targeting the elderly are likely to be smaller and therefore more affordable but cover a smaller share of all the poor. Conversely, many richer countries have older populations, so social pensions are more relevant for addressing poverty as well as more costly; on average, 20 percent of high-income residents are elderly. Nonelderly widows are less than 10 percent of the population in all country income categories on average and less than 5 percent in richer ones.

The strongest correlation between poverty and age is for children, although its strength varies considerably across countries. No countries are in the sweet spot where young children act as an ideal proxy for poverty (the upper right quadrant in figure 5.4). Countries vary significantly in the

Figure 5.4   Relationship of Households with Children Younger Than Age Six Years to Poverty
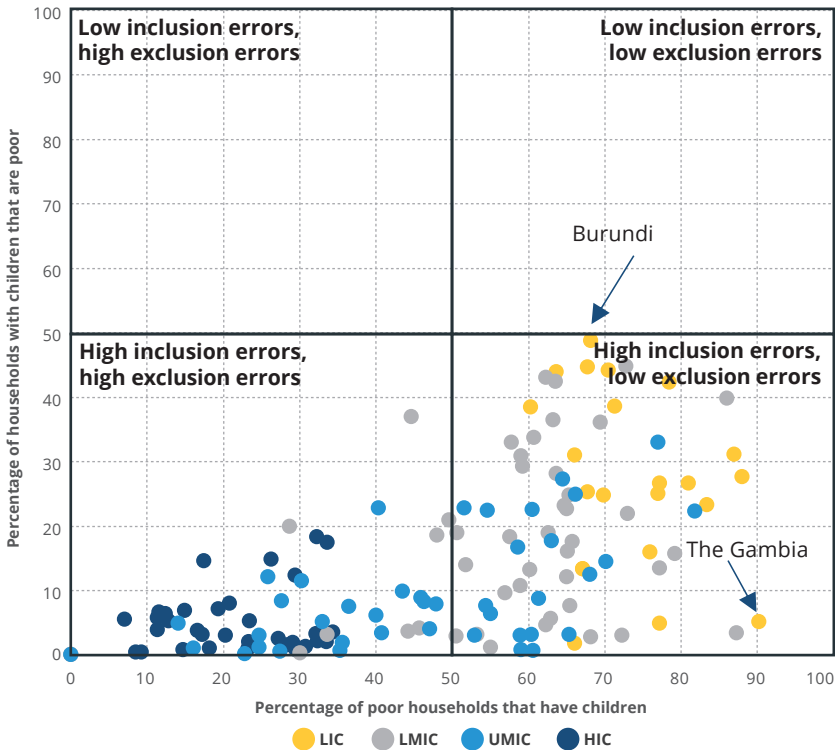
Source: Based on data from the Global Monitoring Database.

Note: The poverty rates use the World Bank's international poverty lines for each income category, which are US$1.90 for low-income countries (LIC), US$3.20 for lower-middle-income countries (LMIC), and US$5.50 for upper-middle-income countries (UMIC). In the absence of an international poverty line for high-income countries (HIC), US$20.00 is used.

percentage of poor households that have young children (shown along the x-axis). In many countries (those in the lower right quadrant), most poor households have children, so child-based programs would exclude relatively few poor households, indeed less than 20 percent in a dozen or so countries. However, in no country is more than half of the households with young children poor, guaranteeing inclusion errors of over 50 percent and also channeling significant budget to nonpoor households. The lower right quadrant contains mostly low-income countries, which are likely to have low tax takes and many competing needs, implying that the cost of covering nonpoor households bears careful consideration. Even within this quadrant, there is significant variation. Ninety percent of the poor households in The Gambia have young children, but only 5 percent of households with children are poor. In contrast, many poor households in Burundi have children (68 percent), while many households with children are also poor (49 percent). Thus, categorical targeting in these two low-income countries would have quite different outcomes. The broader results are yet more varied. For most high-income and half of the upper-middle-income countries, universal coverage of young children would mean both higher exclusion of poor families, as a lower share of them have young children, and higher inclusion error of families with children who are less likely to be poor (lower left quadrant).

Programs for the elderly and widows are less able to do double duty in serving both to reduce general poverty and protect their specific group. Table 5.2 shows that in many countries, the elderly and nonelderly widows make up a relatively small portion of the population. Moreover, the correlation of these categories with poverty is weaker than that of children in many places. As figures 5.5 and 5.6 show for the categories of elderly and nonelderly widows, respectively, almost all countries would have very high inclusion errors and at least as high exclusion errors if such categorical programs were used as the main way to reduce poverty. Even in older high-income countries, less than 50 percent of the poor households have elderly members and less than 40 percent of households with elderly members are poor. Demographically targeted programs for the elderly or widows can be important elements in a wider quilt work or panoply of programs aiming to reduce poverty or manage risks, but they will not themselves have enough coverage to address the problem of poverty in the broader population. In addition, the importance of covering the specific target group itself may vary. Widowhood will have larger economic consequences where women have low education and low labor force participation and are treated unfavorably in family and property laws than where the converse are true. The elderly are more numerous and more likely to live alone in countries further through the demographic transition than in younger countries.

Figure 5.5 Relationship of Households with Elderly Members to Poverty
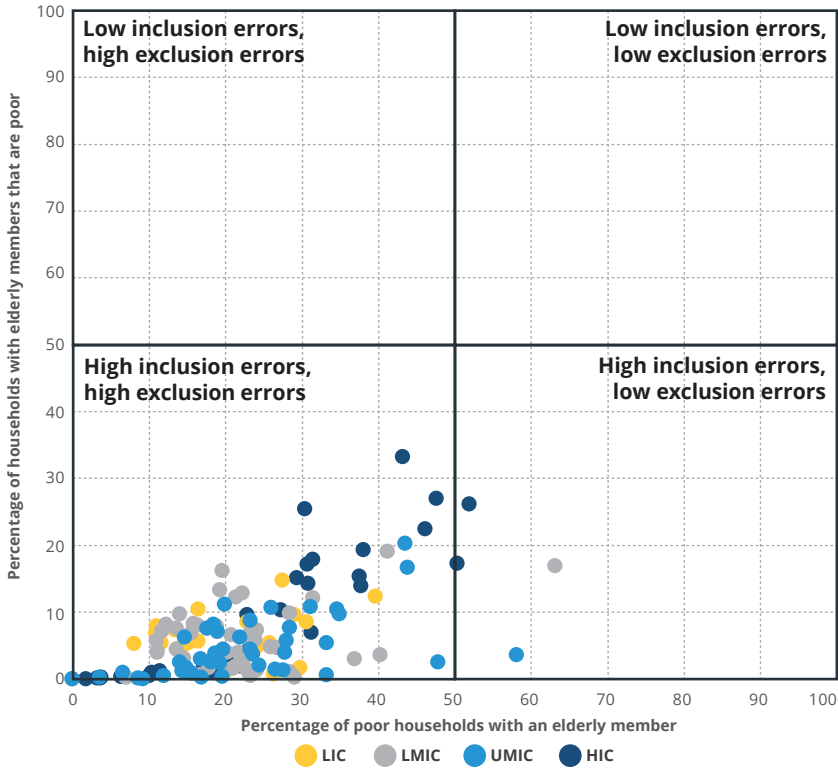
Source: Based on data from the Global Monitoring Database.

Note: The poverty rates use the World Bank's international poverty lines for each income category, which are US$1.90 for low-income countries (LIC), US$3.20 for lower-middle-income countries (LMIC), and US$5.50 for upper-middle-income countries (UMIC). In the absence of an international poverty line for high-income countries (HIC), US$20.00 is used.

Demographic targeting can sometimes play a role in response to shocks, taking advantage of the high coverage many such programs have. The correlation between age and shock will usually not be high—no one's age changes in response to a shock; natural disasters do not strike only those of some ages; and economic disasters hit workers/those of working age more directly and their dependents only indirectly. Nonetheless, top-up benefits to beneficiaries of demographically targeted programs may be a way to get money out quickly, especially where coverage of such programs is high. Of course, children are so biologically vulnerable that it is always important to protect them, and the frail elderly are vulnerable as well. In the case of health

Figure 5.6 Relationship of Households with Widows Younger Than 65 to Poverty
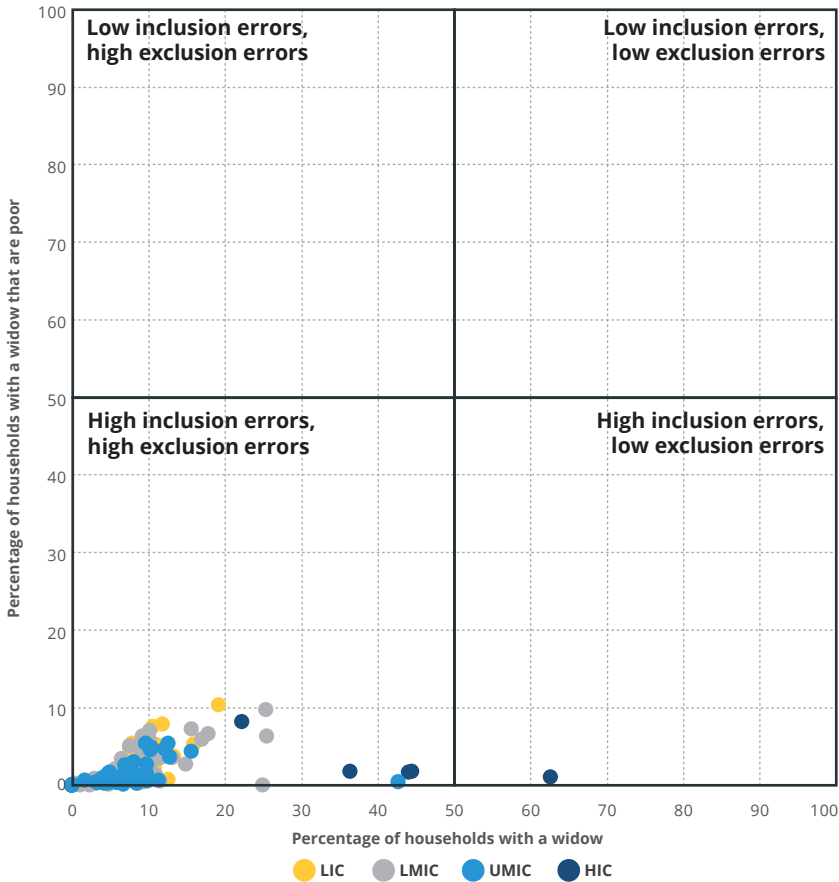
Source: Based on data from the Global Monitoring Database.
Note: The poverty rates use the World Bank's international poverty lines for each income category, which are USUSUS$1.90 for low-income countries (LIC), US$3.20 for lower-middle-income countries (LMIC), and US$5.50 for upper-middle-income countries (UMIC). In the absence of an international poverty line for high-income countries (HIC), US$20.00 is used.

shocks that affect only or disproportionately certain demographics, such as pregnant women (Zika virus), the elderly (COVID-19), and young children (influenza), demographic targeting of access to health care, vaccinations, or cash-based incentives for these can be apt. As shown in box 5.2, following the 2008 food, fuel, and financial crisis, Djibouti targeted income support and nutrition assistance to those who were most vulnerable—poor

households with young women and children—using a combination of geographic, self-selection, and demographic targeting. Box 5.4 describes other cases.

People living with disability or who are facing unemployment are also common categories that may deserve support. However, discerning who has a disability is not simple, nor is determining what level of support may be appropriate for individuals in these groups (see boxes 5.5 and 5.6, respectively). Beyond these short treatments and the references, this book does not cover the topics of disability assessment or labor profiling.

BOX 5.4

Using Categorical Programs in Crisis Response: Examples from Mongolia, Bolivia, and Nepal

As part of its COVID-19 response, Mongolia quintupled payments in its nearly universal child allowance. Before the pandemic, the Child Money Program (CMP) provided an allowance of Tog 20,000 (US$7) per month to children younger than 18 years. It aimed to cover about 85 percent of all children in Mongolia (particularly poorer children). As a part of its COVID-19 relief package, the government of Mongolia increased the benefit amount by five times to Tog 100,000 (US$35), which is equivalent to more than half of the poor's per capita monthly household income. Between April 2020 and July 2021, this cost 4.5 percent of gross domestic product (GDP), one of the larger crisis response programs in the world. It was also the largest element of Mongolia's overall response package, which cost 11 percent of GDP, again among the larger response packages. Data from a series of rapid phone surveys sketch a picture of the outcomes. Overall coverage was high but not universal, reaching about 65 percent of households by the time of the first survey in May 2020 and similar throughout. Coverage was mildly progressive, with 80 percent of households in the poorest quintile receiving the CMP, and declining to 47 percent of those in the highest quintile. Since pandemic-induced income shocks were widespread across the distribution, there were no statistically significant differences in the likelihood of receiving the CMP payment between households that experienced and did not experience a loss in labor income. In terms of usage of the CMP benefits, poorer beneficiaries were more likely to cash out and use the CMP benefits immediately, especially for food and household utilities, while wealthier households were

---

BOX 5.4 *(continued)*

significantly more likely to put the transfer into a savings account for future use (Kim and Uochi 2021).

As part of its COVID-19 Crisis Response Program, Bolivia doubled the benefits of its universal social pension program, Renta Dignidad. Prior to the crisis, Renta Dignidad cost about 1.3 percent of GDP and covered about 91 percent of those ages 60+ and 28 percent of households. The payments represented about 12 percent of total income for eligible households. Starting April 1, 2020, the government doubled the transfer amount for beneficiaries who were not receiving other government pensions. The government also modified the payment mechanism somewhat. It allowed the payments to be made to authorized family members on behalf of beneficiaries, so that the elderly would not have to go to bank branches, and partnered with private banks to increase the number of locations authorized to disburse the transfers. Nonetheless, payments were still in person and only 60 percent of those who were eligible claimed their payments in April. An internet survey in April 2020 captured the situation of families shortly after lockdown orders went into effect and analyzed the effect of the program by comparing the situation of families with a person just shy of eligibility age and those just eligible (an analytic technique called regression discontinuity design). As mobility restrictions were put into effect, across all income levels, 68 percent of households experienced the closure of a family-owned business and 45 percent of households experienced a job loss. Only 52 percent of households reported having enough cash in hand to cover a week's worth of expenses, and only 33 percent of households reported having enough food reserves to cover meals for a week. In addition, 42 percent of households modified their diets and 18 percent experienced hunger. By providing a basic income to beneficiaries, Renta Dignidad led to a 9-percentage-point decline in the probability that someone in the household went hungry. The program also reduced the probability that a respondent reported eating less healthfully and increased the probability that beneficiaries had at least a week's worth of financial resources to cover basic needs. The positive impacts were concentrated among households that were hardest hit by the crisis. They were largest for households with low incomes prior to the pandemic and those in which someone lost their livelihood due to the lockdown policies (Bottan, Hoffmann, and Vera-Cossio 2021).

BOX 5.4 *(continued)*

Nepal used its Social Security Allowances (SSA) as part of its earthquake response. On April 25 and May 12, 2015, two earthquakes struck Nepal, each followed by a series of powerful aftershocks. The earthquakes caused an estimated US$7 billion worth of damage, pushed perhaps three-quarters of a million Nepalese below the US$1.25 international poverty line, and killed 8,831 people. The government of Nepal and the United Nations Children's Fund chose to channel the social protection response via the existing SSA program. The SSA is composed of five cash transfer schemes that use categorical and demographic targeting as the method for targeting.[a] In phase 1 of the earthquake response (May to December 2015), an emergency cash transfer of NPR 3,000 (approximately US$30) was made only to SSA beneficiaries already on the roster in the 19 affected districts.[b] Cash transfers were made to 434,690 individuals, or 93 percent of regular social protection beneficiaries in the selected districts. The speed of the response varied a bit across districts, due to the manual payment process, but most of the beneficiaries had received their payments by October 2015, five months after the first earthquake hit the country. In phase 2 (December 2015 to June 2016), the response focused on the 11 most earthquake-affected districts among the 19 that participated in phase 1.[c] In phase 2, there was a horizontal expansion to serve all children younger than age five years in the selected districts. All children were provided a one-time transfer of NPR 4,000 (approximately US$40). This required opening a registration process to build a roster for all children in this age group in an environment where only half of all children had registered births. The emergency cash transfer program was not envisaged as a replacement for comprehensive humanitarian action, as many other actors were involved in supporting the disaster-affected population (see Merttens, Kukrety, and Majeed 2017).

a. Cash transfers to senior citizens, single women, those with disability, endangered ethnicity, and children younger than age five in selected areas of the country.
b. The 19 districts were Bhaktapur, Chitwan, Dhading, Dolakha, Gorkha, Kathmandu, Kavhrepalanchowk, Kotang, Lalitpur, Lamjung, Makawanpur, Nuwakot, Okhaldhunga, Ramechhap, Rasuwa, Sindhuli, Sindhupalchowk, Solukhumbu, and Tanahun.
c. The 11 districts were Sindhupalchowk, Nuwakot, Dhading, Gorkha, Rasuwa, Kavhrepalanchowk, Dolakha, Sindhuli, Ramechhap, Makawanpur, and Okhaldhunga.

BOX 5.5

Challenges and Considerations around Disability as a Category in Determining Eligibility or Benefits

Disability is complicated first and foremost because of the range and depth of impacts it can have on all sorts of aspects of the lives of people who live with one or more disabilities, but also because of the social policy initiatives meant to improve their capabilities.

The Convention on the Rights of Persons with Disabilities (UN 2006) recognizes persons with disabilities as individuals "who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others."[a] The degree to which an impairment affects the ability to carry out daily activities and work depends not only on the severity of the impairment, but also the demands of the environment and specific activities. Using a wheelchair is a much larger barrier when the built environment is full of stairs and narrow doorways than when universal design elements are ubiquitous. The demands of the workplace can matter powerfully as well—a person with a blurring in the center of their visual field will find the impairment a greater barrier to work in needlecraft or neurosurgery than to work in modeling or psychotherapy.

The full agenda for disability inclusion is many faceted: prevention and rehabilitation to minimize impairments; accessibility in the built environment, transportation, and information and communications technology to minimize the barriers for those with impairments; inclusive education and training to maximize the skill sets of people living with disability; societal and employer attitudes attuned to abilities and problem solving about barriers; and so forth.

Income support via social assistance is important in two ways. When efforts toward disability inclusion have been inadequate to result in employment or earnings sufficient for independence for the person with a disability, then income support may take the form of income replacement. Even when barriers to employment have been addressed, social assistance can be an important income supplement to help mitigate the extra costs of living with disability, for example, for medical care and perhaps assistive devices, extra costs for mobility, sometimes the expenses (explicit or implicit) of caregivers, and so forth.[b] While calibrating support to help move people into work will need to be customized to the individual and context, benefits for the extra costs of

continued next page

BOX 5.5 *(continued)*

basic life might be thought of as more categorical, but a disability assessment is needed.

A disability assessment is a field of endeavor at least as complex as that of quantifying the economic dimensions of welfare, with equal concerns about errors of exclusion and inclusion. Thus, it is the subject of other books and studies. WHO and World Bank (2011) cover the standards for the International Classification of Functioning of Disability and Health (WHO 2001). Bickenbach et al. (2015) discuss how the standards are moving into social protection policy globally. World Bank (2020a, 2020b) provides case studies of Latvia and Lithuania. United Nations (2015) presents the report of the Special Rapportuer on the Rights of Persons with Disability and Social Protection.

The World Health Survey estimates a prevalence of disability in 2004 of about 15 percent, ranging from 12 percent in higher income countries to 18 percent in lower income countries. This figure refers to adults who experienced significant functioning difficulties in their everyday lives. The prevalence of disability in lower income countries among people ages 60 years and older, for instance, was 43 percent, compared with 30 percent in higher income countries. The average prevalence rate for adults with very significant difficulties was estimated at 2.2 percent. The Global Burden of Disease studies come out in the same ballpark. Mitra et al. (2021) look at 21 low- and middle-income countries with information collected using the Washington Group Short Set questions. They find that the median prevalence stands at 10 percent among adults ages 15 and older, and that 23 percent of households have a member living with disability. It is generally agreed that prevalence is higher among the poor because poverty exposes people to many risks that can result in disability and people with disabilities face barriers to employment (WHO and World Bank 2011). Moreover, not allowing for the extra costs of disability can underestimate poverty among those with disabilities and for a nation as a whole. Zaidi and Burchardt (2005) estimate that taking such costs into account would raise the poverty rate among pensioners in the United Kingdom by 18 percentage points and overall by 3 percentage points.

a. UN Convention on the Rights of Persons with Disabilities, article 1, https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-1-purpose.html.
b. Estimating the extra costs of disability is a nascent field. Methods, data, and definitions differ, and studies are rare, especially in developing countries. Mitra et al. (2017) provide a good review. Estimates of such differential costs in lower-middle-income countries are in the range of 8 percent in China, 9–12 percent in Vietnam, and 14 percent in Bosnia and generally much higher in Ireland, Spain, the United Kingdom, and the United States, on the order of 11–70 percent, depending on the degree of disability and family configuration.

BOX 5.6

Profiling Job Seekers to Differentiate Support

An important part of social protection is the support that helps people get jobs. There may be different groups of focus: youth, as new entrants to the labor force; those (most often women) returning to the labor force after a period of caregiving; or people facing unemployment, especially those displaced from industries affected by structural changes that diminish demand for the whole sector. A special concern can be avoiding long-term spells of unemployment or withdrawal from the labor force as these can result in scarring, with lower chances of (re-)employment and/or lower wages.

Countries with well-developed, active labor market policies, usually those with predominantly formal labor markets, typically offer a range of supports of different intensities. This can range from self-service access to job listings; to job search assistance and coaching; to skills training in any or all of a variety of basic literacy and numeracy, vocational, or socioemotional skills; to wage subsidies or temporary public employment; to sheltered employment for people with severe trauma or disability; or some combination of these policies. Use of some of these services may be paired with or even a condition for income support via unemployment insurance or social assistance.

Matching the more intensive and costly of these services with the people who would be most likely to benefit from them is a targeting problem that is somewhat analogous to providing greater levels of income support to the relatively poor.[a] Profiling job seekers has the same problems of errors of exclusion (where a person who receives too few supports may not find employment) and inclusion (where a person who might easily find employment receives higher levels of support and uses more resources than needed).

There is a range of profiling techniques to try to make good matches between the risk of long-term unemployment and services to avoid that outcome. Some profiling methods rely more on simple rules (for example, on age or duration of unemployment or job-seeking spell), some rely more on the human skills and judgments of social workers, some rely on data and modeling to predict which job seekers will need the most supports, and some use a combination of these approaches. There is a literature on the pros and cons of these different methods that is akin to the choice of household-specific methods for assessing monetary welfare. The statistical profiling

continued next page

BOX 5.6 *(continued)*

literature in particular has a good deal in common with the proxy means testing literature about which data and methods of inference will be most suitable. In general, the profiling defines the treatment track (for example, whether or how much training support will be provided), but it does not get into granular detail, over which the individual and case worker maintain some agency (for example, the specific training program in which to enroll).

Useful references include Barnes et al. (2015); Desiere, Langenbucher, and Struyven (2019); and Loxha and Morgandi (2014).

Source: van Landeghem, Desiere, and Struyven 2021.
a. The analogy is not perfect, as the most intensive active labor market policies may be offered to those at medium risk of long-term unemployment, with those at low risk expected to gain work with little support, and those at highest risk of unemployment referred to other social services to handle issues such as health and rehabilitation, substance abuse, and so forth, with active labor market policies available but not the first focus.

Methods That Rank People According to Welfare

The choice among household-specific targeting methods (when they are to be used) depends on several factors. Factors related to the availability of information are well known, but there are others as well—related to the institutional landscape and social contract.

Factors Related to Formality and Information Sources

The availability of administrative data is a traditional consideration in choosing among methods for household-specific assessments of money-metric welfare.[17] The following are particularly influential: (1) the share of formal employment and formal earnings; (2) the capacity of the administration to observe and verify the income (and assets) of the country's population through databases designed for other purposes—income tax, social security contributions, land registries, car ownership, passport use, payment of utility bills to state-owned enterprises, and so forth; (3) the quality and frequency of such data; and (4) the ease with which data from different sources can be matched.

When the administration can easily verify the main sources of income and assets of the applicant population, the means testing method can be accurate and is often viewed as the gold standard as it is the only method that does not include any inherent errors. This method is typically found in

Organisation for Economic Co-operation and Development/high-income countries, with their highly formal economies and extensive data systems that allow for verification. There are also applications in middle-income countries, such as Brazil, China, and South Africa, which started with little verification of declared income. Over time, among other actions, such as improved interoperability, these countries have built robust monitoring and evaluation systems to identify inconsistencies and effective strategies to communicate to the population the consequences of false declaration, which would lead households to suspension of benefits and penalties that would block any member of the family from receiving any type of assistance. Means tests stand alone as the only method used in half the instances in which they are used. Means testing is also commonly used to restrict categorical programs, such as child allowances or social pensions for needier families; when means testing is used with another method, 83 percent of the time it includes categorical eligibility.

The definition of income used for eligibility purposes—sometimes referred to as the administrative or program definition of income—may differ slightly from the economic definition of income. Some types of income that add welfare may not be included in the administrative definition, such as transfer income from other social protection programs or portions of earnings that may be disregarded to avoid an income trap that would discourage work or run counter to the objective of the program of focus. Other types of income—typically occasional, rare events, or those that cannot be verified—may also be disregarded. However, the two measures, administrative and economic definitions of income, will be closely correlated in terms of ranking as well as levels.

When formality is insufficient for a verified means test, an HMT that relies on the measurement of formal incomes (observable and verifiable) and the estimation of some of the informal incomes to produce an estimate of total welfare may be appropriate. An informal employment rate of over 20 percent will reduce the precision of a means test. However, if formal employment is somewhere between 50 and 80 percent, the economy has a large share of its income in the formal sector. Hence, income is "visible" to the administration, tilting the balance toward methods of observing rather than estimating the welfare of applicant households. Asset registries—of land ownership, dwellings, cars, and so forth—can be used to impute incomes, to add to the verifiable sources or as filters for affluence testing.

When a large proportion of jobs are in the informal sector, it is difficult to observe, measure, and/or verify the level of income or consumption and assets of a household. Therefore, it is difficult to identify the population using observed welfare measures, which calls for an alternative approach.

This was the motivation when the PMT was invented in Chile 40 years ago; it is also why the method caught on and spread among the many other countries with significant informality.

The PMT method consists of estimating household welfare based on a set of relatively easy-to-observe indicators such as individual demographic characteristics, aspects of the dwelling, ownership of household durable goods, or location. The parameters of PMT programs are determined using a representative household survey that collects information on household welfare and the characteristics correlated with it and statistical modeling. The precision of the PMT depends on the quality of the information, the procedure used to estimate the coefficients, and the strength of the association. However, because it uses inference rather than measurement, PMT involves inherent error.

PMT is generally combined with another targeting method. In over three-quarters of the ASPIRE observations where PMT is used, it is part of a mixed methods approach. Categorical (69 percent) and geographic (48 percent) methods are the most common partners, but PMT commonly appears with all the other methods, including community-based ones (31 percent) and means testing (23 percent).[18]

In any of means testing, HMT, or PMT, the calculated welfare measure is compared with an eligibility threshold, and different thresholds may be used for different programs. The threshold(s) can be estimated based on administrative data, for example, if the country has a comprehensive income tax, or from representative income surveys to get an idea of how many people would qualify at each level. For PMT, the use of multiple thresholds presents additional complexity in statistical modeling of the relationship between proxies and true welfare, a theme taken up in chapters 6 and 8.

As a country becomes more formalized and improves the interoperability of its databases, it may progress from PMT to HMT to means testing, as each later stage requires greater capacity but is more accurate. The case study of Chile in box 5.7 is a classic example of this progression. It was the first country to deploy a rudimentary PMT in 1979. As program objectives, data availability and technical capacity, evolved, it moved to a version of HMT and now uses means testing. Several countries that use PMT, such as Albania, Armenia, Costa Rica, Georgia, and Turkey, include formal sector information (on formal wages, social protection program benefit, or receipt) in their PMT formula or as exclusionary filters. As part of its regular review of the performance of its scoring systems, Armenia has examined moving from PMT to HMT, but no decision has been taken at the time of writing. Saudi Arabia is considering transitioning to an HMT.

BOX 5.7

Chile: Evolution of the Targeting System in a High-Capacity Country[a]

For many years, Chile used proxy means testing (PMT) as the main targeting method for poverty-targeted social assistance programs. The eligibility form for the scoring system was Ficha CAS. The first version of the Ficha CAS was developed in 1979 using principal component analysis. Households were classified using a discrete score ranging from 1 to 5, and the score was used to allocate the crisis response cash transfer and workfare programs. In 1991, the eligibility form was expanded and applied during home visits, allowing the verification of housing conditions. A new formula scored households on a continuous PMT index that estimated household income. A two-year recertification period was also introduced. This continuous estimate of household income was used across a larger number of programs, with different eligibility thresholds and sizes. In 2007, the data collection instrument changed to Ficha de Protección Social, and the PMT formula changed from predicting household income to estimating income generation capacity and incorporated some features of a hybrid means test (HMT). It also used the national identification (ID) to validate identity and cross-verify the data with pension databases.

Between 2007 and 2012, the Ficha de Protección Social was applied to an increasing share of the population, from 5 million to 12 million people, and was used to allocate resources for 80 different programs and benefits. This expansion put pressure on the social protection system as many programs did not have resources to cover all the people meeting the eligibility criteria, resulting in waiting lists, as well as leading people to request out-of-cycle recertification (and lower HMT scores through distortion of self-declared information).

In 2014, the Chilean government decided to implement a new targeting system, the Registro Social de Hogares (RSH), which ranks households in seven groups based on an income and assets test (means test). For the first time in Chile's history, the RSH was allowed to collect information on income and assets on behalf of applicants from other administrative databases, based on their informed consent and with regulations for respecting personal data protection. Thus, the RSH has not embarked on collecting new information from applicants; it only accesses and uses existing information from administrative databases (tax records, wages, social security contributions, health insurance [public and private] contributions, unemployment insurance,

continued next page

BOX 5.7 *(continued)*

pensions [contributory and noncontributory], education records,[b] real estate, and vehicles).

The gradual development of the targeting instrument in Chile was possible due to political commitment to reform, a desire to improve coordination across stakeholders, and investments in the delivery system's human and physical capital. The reform resulted in better data integration and system interoperability, efficient use of existing administrative data, and cross-verification of self-declared information. This degree of integration and interoperability was only possible due to the large coverage and use of the national ID system (Rol Único Nacional) throughout the country. The use of a unified and single assessment of household means ensures horizontal equity across multiple programs. According to the government, this improvement has led to a reduction in inclusion errors and a better public understanding of selection criteria.

The Chilean case emphasizes the importance of continuously monitoring learning and improving over time. Each change was built on the experience accumulated in previous phases. The targeting system began and evolved to provide support for social policies and adjusted over time to the objectives of such policies. The targeting method has been modified consistently with a prioritized approach to social policies and programs, not by a particular program or development partner. The system has benefited from advances in technology, but also invested significantly in communication and delivery systems so that federal and municipal bodies could contribute to the social protection system.

Sources: Clert and Wodon 2001; Larrañaga 2005.
a. Ficha CAS, 1979–90; Ficha CAS 2, 1991–2006; Ficha de Protección Social, 2007–15; and Registro Social de Hogares, since January 2016. For more details, see Silva et al. (2018).
b. From the education level of individuals and school enrollment database.

Means testing and HMT are both suitable for shock response if changes in well-being are quickly captured by the information systems used. Changes in formal income may be reported at short intervals through the social security contribution records, for example, but only annually and with delay in full income tax records. Ownership thresholds of or imputations from assets may have to be altered for natural disasters, as land or property registers are unlikely to be updated quickly to reflect damages or the likely loss of earnings from weather-related disasters. In general, as they rely heavily on good data integration, most shock responses using these

methods are fast given that people just need to trigger benefits by using the existing application process.

PMT by nature predicts long-run welfare and, in practice, registries using PMTs do not usually update information frequently.[19] These features make PMT too static to be as effective in responding to shocks, at least without adjustments. This includes both idiosyncratic shocks where household assets, housing, and demographic characteristics do not change significantly when sickness or job loss occurs, as well as broader economic shocks. However, adjustments that involve using information about the risk of weather-related shocks as part of PMT can develop triggers for vertical or horizontal expansion of programs following such shocks; chapter 6 discusses these issues, including the PMTplus method. However, PMT may be serviceable for ranking households for programs meant to address chronic poverty.

CBT may be the conceptually preferred method when governments wish to devolve the definition of welfare to communities. It may be practically preferred when it seems that communities have better information than official administrators have (from existing databases) or could practically gather (for example, in a survey sweep or application process). The philosophy of information gathering is very different for a CBT-based method. It does not rely on databases or statistics, which may be absent or unreliable, but on the tacit knowledge of neighbors who observe markers of each other's welfare in the course of daily life. The first minimum condition that would need to be met would be that community members know each other well enough to rank or assess each household's needs. In Niger, Premand and Schnitzer (2018) found that in many communities, members were unable to rank some other households, although by using multiple committees, all households could be ranked. Moreover, when the number of households to be ranked is large, fatigue sets in. In Indonesia, error rates were between 5 and 10 percentage points lower for the first household than for households ranked in the latter half of the meeting (Alatas et al. 2012). In Djibouti, CBT was used for targeting in rural areas, but PMT was used for urban areas on the basis that community cohesion was not as high in urban areas (see box 5.2).

There are sometimes tensions between the rules established by the central agency financing the program and how communities or their elites wish to implement them. Sometimes communities are given some guidance on the notion of poverty the program administrators hold, but the exact information considered and the weighting of different factors are only implicit and presumably may vary from community to community even within a country. For example, in Malawi, community committees would exclude some households from programs because they were already receiving other assistance (considered double dipping) even if this undermined

the program's objectives (Lindert et al. 2018). Communities sometimes share benefits, diluting their effect, as was done with Indonesia's Raskin subsidized rice program until it was phased out (World Bank 2010b); or, out of jealousy, communities may implicitly or explicitly demand a share from beneficiaries, as in Chad (Della Guardia, Lake, and Schnitzer, forthcoming). Alternatively, community committees may come under pressure from elites to include their family members, although there are various ways to minimize this in the design of community processes.[20]

The effectiveness of CBT in the aftermath of a shock depends on how quickly communities can reassess their members' changed needs. As committees must meet to identify the cases, response times will be in the range of other postdisaster assessment methods. Communities are also known to consider individual household circumstances when targeting in normal times, such as whether someone had recently suffered an illness or accident or lost a job. Such local knowledge can help overcome the lack of formal data on idiosyncratic shocks in less developed contexts.

Factors Related to Administrative Systems

Targeting capacity must be built, and the sort of effort required varies among the methods. To build means testing capacity where none has previously existed usually involves developing client interfaces, which tend to be continuously available through a mix of in-person and virtual means. Building targeting capacity may require work on the databases used in verification to improve them and make them more compatible (this work may improve not only the capacity to target, but also the other functions of government that the data were meant to support, especially development of direct taxation on income and assets). HMT involves the same, plus some modeling to figure out the estimation techniques used in the parts of income to be estimated. PMT similarly requires modeling work upfront. The administrative capacity it requires depends on whether a survey sweep, an on-demand approach, or a mix of both will be used for client interface. The periodic survey sweep approach generates a "lumpy" rather than continuous staffing need, which has implications for budget—this sort of spike in cost is amenable to project financing from development partners, but it can be harder for domestically financed programs to handle. Community-based approaches also require some field capacity to animate and supervise the community processes, which may also be done only periodically and thus may be lumpy.

The degree of interface with other processes, programs, and agencies is correlated with the targeting method. A permanent cadre of intake officers can be used in that sole function and for a single program. However, often the same officers provide intake or referral services

across a range of programs and/or provide functions beyond intake, such as onboarding, general information queries, a touchpoint for grievances, referrals to other services, and sometimes even providing counseling for clients where the program provides for job search assistance, psychosocial support, and so forth. These joint functions can provide a more convenient service to clients, lowering their transaction costs and increasing their trust in the state. A contracted-out survey field worker for a PMT, in contrast, would not be expected to provide information or functions beyond the collection of data. If the PMT targeted program does not have other local staffing, people do not have an obvious point of contact to answer questions about their eligibility, program rules, or any hitches during receipt of benefits. In CBT, practice can vary. The aspiration may be that the community representatives learn enough about the programs to help with outreach and to help community members know how to address any questions or complaints they have, although accomplishing this takes a good deal of training, which has not always been done. In PMT, the modeling and information technology (IT) capacities can be independent from other agencies. This approach may be viable when data sharing is not allowed between agencies and there is no political will to solve that.

The nature and frequency of grievances and their handling will depend on the targeting method. Means tests may be conceptually easy to explain, but they often have complex formulae and may rely on data cross-matches or verifications that lead to errors that may be the subject of grievances. PMT and to a lesser degree HMT have inbuilt statistical error. Even with full, accurate information that is properly reported and handled in all processes, a poor person can be judged not eligible. It would be unsurprising if that person wanted to file an appeal and have their case reconsidered. Programs need at minimum to provide some grievance procedure where the person can at least verify that their information was handled correctly and that there were no clerical mistakes. This will still leave such persons uncovered and the community around them with lower confidence in the accuracy of PMT or HMT, but it is at least the minimum standard of recourse that should be expected in a means test.

Some countries have processes that allow for some overriding of the targeting formula to mitigate the statistical error of PMT or HMT. A community validation phase is fairly common, but the statistics are not well compiled. Paes-Sousa, Regalia, and Stampini (2013) recount how community validation has been used in several Latin American programs, but it has been discontinued in some as it was considered relatively marginal in importance. It may be more helpful in bringing in households missed in enumeration than challenging errors of inclusion (Jones, Vargas, and Villar 2008). Several African programs still maintain the community validation

phase, for example, Kenya's Orphans and Vulnerable Children program, Tanzania's Productive Social Safety Net, and Malawi's Social Cash Transfer Programme (Arruda 2018). In the Republic of Congo's Lisungi program, the grievance redress mechanism handles complaints related to inclusion or exclusion errors. Agents of the Agency for Counter-Verification may overrule the local committee's initial decision or the PMT in case of inclusion or exclusion of households after the claim treatment process is completed. A few programs have or have had systematic and a bit more institutional systems to allow judgment-based overrides of the PMT formula. In Armenia, local social protection councils (consisting of five local government representatives responsible for the social sectors and five representatives of locally active nongovernmental organizations) are given the authority/discretion to allocate 3 percent of the actual expenses paid to Family Benefit Program beneficiaries in a given district to applicants who do not qualify for the benefit according to their vulnerability score but are found to be needy based on the results of social workers' home visits and other considerations.

Privacy is an important principle in the human rights framework and harder to respect for some targeting methods than others. The principle is that personal information should be kept private, and that data should be collected with the knowledge and consent of the subject, accessible to him or her, accurate, complete, and up-to-date. Some parts of targeting practice do not respect this standard well.

- It is not clear that CBT would meet consent standards since the data used come from neighbors having observed people in their daily lives, without formal prior consent. For the other methods, applicants can be asked to sign a consent form prior to conducting an interview or as part of an online application process, although realistically, those badly in need of income support may feel that denying consent is not a real option.
- The community validation of lists of beneficiaries determined through CBT or PMT—often thought to be important in remedying errors of inclusion and exclusion—would seem to violate the right to privacy. So would the practice of public posting of beneficiary lists determined through any method. This is sometimes done as a means of transparency and sometimes as a practicality to communicate to all considered at low transaction cost, which are both positive objectives but in opposition to privacy.
- Survey sweeps only every few years—adopted principally to lower administrative costs and as a way of providing outreach—violate the accurate and up-to-date standard unless there is a way to get on-demand reassessment between sweeps.

- As noted in chapter 4, privacy and security protections on databases used or built for eligibility determination are necessary, and they are a focus of policy action in countries that are working to secure the right to privacy.

Transparency and social cohesion are other desirable attributes that are harder for household-specific methods than for categorical ones. With means testing or HMT, the notion that the method is supposed to sort the poorer from the richer is usually understood, but the details of what counts in the calculations may be less so. PMT may be understood by policy makers, but it is often not well understood by communities. Accounts of the acceptability of CBT differ. All these methods draw lines within communities in ways that risk "othering" or social tensions. These disadvantages must be set against the method's power to rank or assess welfare more accurately than the non-household-specific methods.

Addition of Quantitative Information to the Decision-Making Process

Although much of the decision on which targeting method to choose will be based on the qualitative factors just discussed, some empirics can also be helpful. One part of these comes from country-specific analysis, simulations, and modeling, and another part comes from international evidence.

Country-specific simulations can help quantify for a specific country some of the trade-offs involved in different design parameters—definitions of eligibility and targeting methods, the level of benefits, and the cost of the program—via their impacts on poverty or measures of targeting outcomes. A common scenario is to simulate different targeting methods for a fixed budget: a program that gives benefits to all children or to all the elderly versus a geographically targeted program, versus one that ranks households with a means test or PMT, or one that combines some or each of these methods. Often perfect targeting and universal benefits are simulated as well, not so much as policy proposals but to anchor the endpoints of the spectrum of choices. A complementary set of simulations may be done with the same scenarios for targeting methods but with the benefit kept constant and the required budget allowed to vary.

The simulations will be only approximations and must be interpreted as such. Simulations usually assume perfect implementation (100 percent take-up and no errors or fraud). Sometimes analysts try to make some allowance for imperfect execution, for example, by assuming that a certain portion of the target group will fail to apply for the program or be misassessed. Analysts may also reduce the benefits to be distributed by different amounts to approximate administrative costs. More often, the scenarios

are left simple and the caveats are handled in interpretation. Even with adjustments, the simulations cannot deal so well with intangibles, for example, the social acceptance of some methods over others or the interaction between methods, politics, and budgets.

The metrics used to assess the simulations are important:

- This chapter recommends giving great weight to the impact on the poverty gap in the assessment. This is relatively simple to interpret and explain to policy makers and the public. It embodies a social welfare function that gives greater weight to the poorest than the less poor and the only just poor. It will thus register changes when a transfer is received by a poor person, even if the transfer is insufficient to lift that person fully out of poverty.[21]
- Understanding the impact on the poverty headcount is also useful and easy to explain. However, it gives a lot of weight to transfers to those just around the poverty line rather than to the poorest as the former are easier to bring out of poverty.
- Looking at errors of exclusion or inclusion separately may also be useful as there may be special political sensitivities to these, but they are inherently partial measures.

An example of such a simulation is presented in table 5.3 and figures 5.7 and 5.8, comparing various categories with a household-specific method and pure geographic targeting. The simulations are for a large and diverse middle-income country with average levels of inequality and poverty at the $3.20 line of around 30 percent. In the example, a policy maker who wants to reduce poverty has a budget of 0.5 percent of GDP. With a 30 percent poverty rate, this budget implies an average transfer equivalent to 14 percent of the poverty line for each member of the household.

The policy maker then has her analyst simulate different program designs and approaches. She simulates beneficiary eligibility based on:

- A household-specific method. In this country, the policy maker may judge, based on rates of formality and the coverage of administrative registers, that PMT is the most pertinent among means testing/HMT/PMT. Of course, the same sort of comparison can be done simulating the outcomes of means testing or HMT if one of them is deemed more pertinent, as shown in figure 3.3, in chapter 3, for HMT and box 5.8 for means testing. It is harder to simulate CBT. For guidance between PMT and CBT, the next subsection reviews the evidence from field experiments with comparative treatment arms.
- Geographic targeting.
- Categorical: all households with children younger than six years.
- Categorical: all households with elderly members over age 64.
- Categorical: all households with a widow younger than age 65.

Table 5.3 shows how much of the population is covered under each approach and the benefit levels given the fixed budget. Since 43 percent of the population lives in a household with a young child, the benefits are diluted to 10 percent of the poverty line. Conversely, since only 17 percent of people live with an elderly household member and 9 percent with a widow, the benefit levels increase to 25 and 45 percent of the poverty line, respectively, to exhaust the fixed budget. The higher benefit level for households with elderly members makes sense as the payments can be considered as replacing income rather than supplementing it; the even higher widow benefits might be scaled back and the budget reduced. With PMT, the households with the lowest scores receive benefits until the budget is exhausted; for geographic targeting, all households in a poor area receive benefits, and the program continues to cover new areas until the budget is met, which means that people in 185 of the 500 districts are covered. Both approaches cover 30 percent of the population.

In addition, the analyst looks at mixed methods and tiered benefits. First, because PMT scores vary by household, benefits can be tiered depending on how low the score is; two benefit levels of 29 and 12 percent of the poverty line are used. Second, the lower benefits if a universal young child grant is made can be increased to the same level as the other approaches if a rationing device is used. Two mixed methods approaches seek to do this. One applies PMT scoring to all households with young children. The other provides benefits to all households with young children in the poorest districts until the budget is exhausted (which ends up covering 367 districts). Both approaches cover 30 percent of the population with a benefit equal to 14 percent of the poverty line. The three remaining approaches are thus:

- PMT with tiered benefits
- PMT with children younger than age six
- Geographic targeting with children younger than age six.

Table 5.3   Program Coverage and Benefit Levels under Different Eligibility Approaches

	PMT	PMT with tiered benefits	GT	HH with a child < 6	Elderly	Non-elderly widows	PMT with a child < 6	GT with a child < 6
Coverage of population (%)	30	30	30	43	17	9	30	30
Benefit relative to poverty line (%)	14	*	14	10	25	45	14	14

Source: Original calculations for this publication.

Note: * means varies by tier, 29 for lower tier, 12 for higher tier. GT = geographic targeting; HH = households; PMT = proxy means testing.

The analysis considers the coverage of the different methods across the income or consumption distribution. Figure 5.7 shows what percentage of each decile is covered by each eligibility approach. All the methods exclude some poor households and include some rich households. PMT does the best, covering 79 percent of the poorest decile, 60 percent of the next poorest, and 9 percent or less of the richest three deciles. When PMT is applied only to households with young children, it performs less well. Only 58 percent of the poorest decile is covered, reflecting that this approach automatically excludes poor households without young children, while coverage of the second richest quintile is nearly 20 percent. Geographic and geographic for households with young children targeting perform similarly and slightly less well than PMT for households with young children. The pure categorical approaches reflect the different population representations of their categories. Covering all households with children covers 60 percent of the poorest decile but much higher rates of all other deciles than the other approaches as well, including nearly 30 percent of the richest decile. The elderly- and widow-based approaches cover less than 25 percent of all deciles, and the coverage for households with widows is almost flat across the distribution.

Figure 5.7   Simulated Program Coverage under Different Methods, by Decile
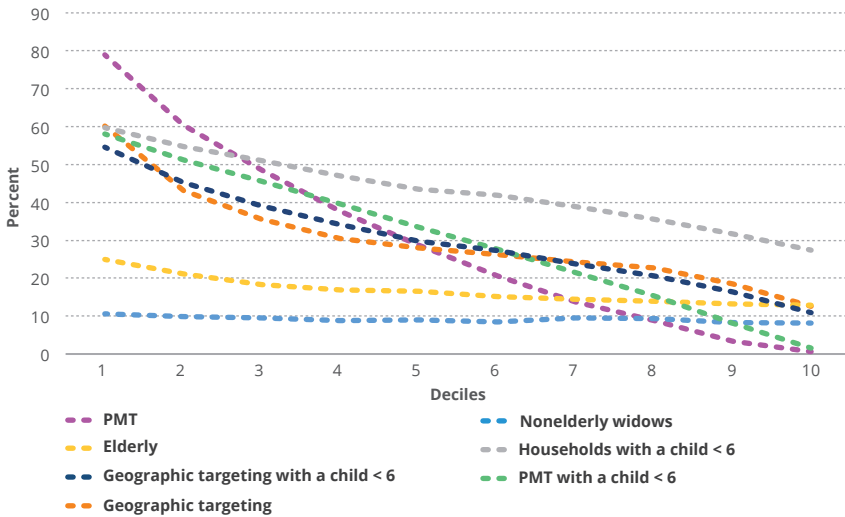
Source: Based on data from the Global Monitoring Database.
Note: PMT = proxy means testing.

The proportion of total benefits going to poorer households can also be simulated. With PMT's greater coverage of the poor and less of the rich, 45 percent of all benefits goes to the poorest two deciles (figure 5.8) and less than 16 percent goes to the richer half of the people. Tiering the benefits with PMT results in 54 percent going to the poorest two deciles. PMT on households with young children would send 36 percent of benefits to the poorest two deciles and 25 percent to the richest half. Geographic and geographic with young children see these two groups receive nearly equal benefits: 34 percent to the poorest two deciles and 35 percent to the richest half for straight geographic and 33 percent each for geographic with young children. The purely categorical approaches result in considerably more benefits going to the richest half than the poorest two deciles: 41 and 27 percent, respectively, for households with young children; 42 and 28 percent, respectively, for households with elderly members; and 48 and 22 percent, respectively, for households with widows. The simulations

Figure 5.8 **Simulated Program Incidence under Different Targeting Methods**
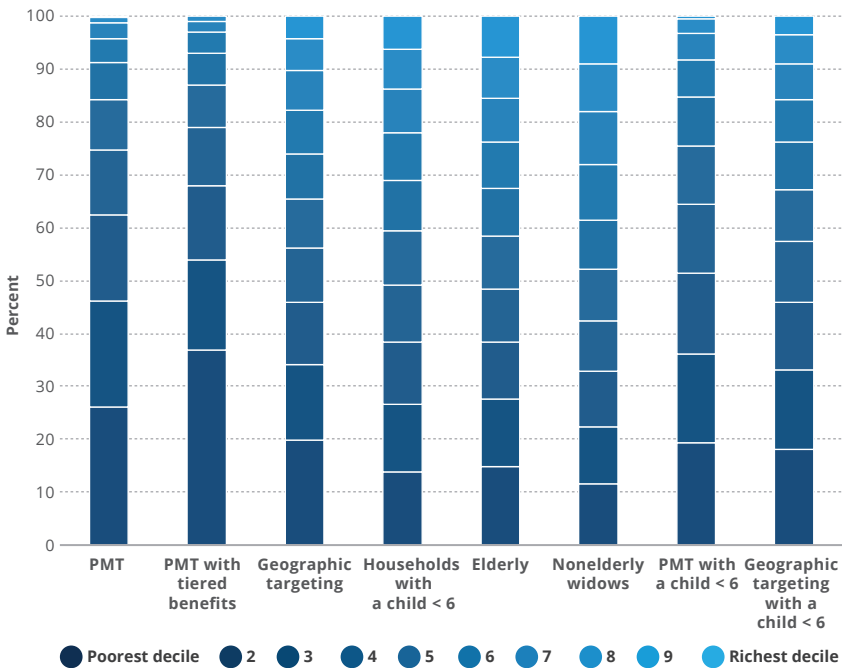
Source: Based on data from the Global Monitoring Database.
Note: PMT = proxy means testing.

suggest that compared with the other methods, a PMT approach will send a greater proportion of benefits to the poor, while the categorical approaches leak much of the budget to richer households that have members of the favored "vulnerable" groups.

The analysis also shows how many poor people are excluded or nonpoor included. When the number of beneficiaries is the same as the number of the population intended to be covered, inclusion and exclusion errors are equal. That is, if a poor person is missed, that means that a nonpoor person has been mistakenly included. The errors in figure 5.9 reflect this. PMT has the lowest errors, albeit nontrivial, at 37 percent; errors for PMT for households with young children rise to 48 percent and for geographic targeting to 54 percent. For categorical programs where the coverage is different from the size of the intended population, the inclusion and exclusion errors are not the same. The larger child program—covering 43 percent of the population compared with the poverty rate 30 percent—means that the exclusion errors are lower (at 43 percent) than the inclusion errors (at 62 percent). This occurs because more households are covered, so there is less chance of excluding the poor but more chance of including the non-poor. For the smaller programs for households with elderly members or

Figure 5.9   Inclusion and Exclusion Errors
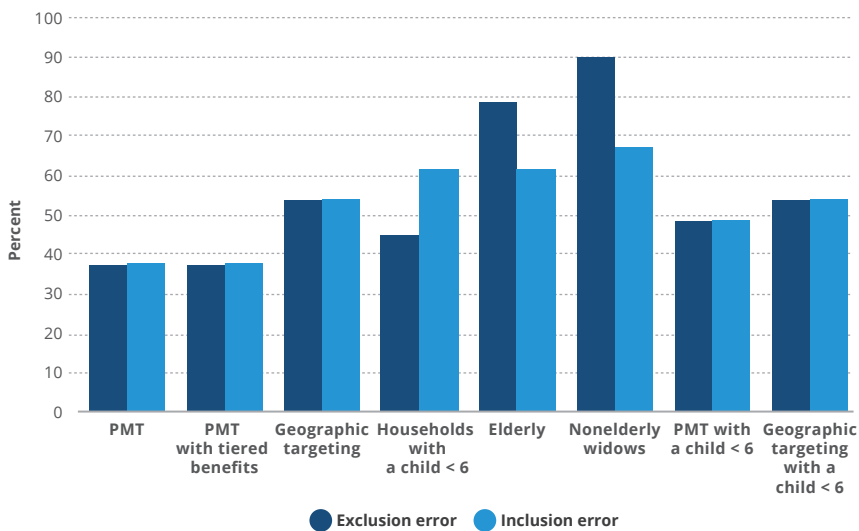
Source: Based on data from the Global Monitoring Database.
Note: PMT = proxy means testing.

widows, the opposite result occurs—high exclusion errors (79 and 90 percent, respectively) since so many poor households do not include members in these categories, yet still high inclusion errors (62 and 67 percent, respectively) because many households with this demographic composition are not poor.

Finally, the impact on poverty can also be estimated. The policy maker's initial objective was to reduce poverty. The simulation estimates the extent of reduction of the poverty rate and the poverty gap.[22] In the analysis, the PMT approach reduces the poverty rate the most (by 3.6 percentage points) as well as the gap (by 2.0 points) (figure 5.10). Using the PMT score to tier benefits would reduce the poverty rate less (2.9 points) but the gap more (2.8 points); the higher transfers to the poorest are not enough to pull them over the poverty line but go further to raise their incomes. The geographical and categorical approaches do a similar job at reducing the number of poor as the tiered PMT approach, but they reduce the gap by less, while the mixed method use of PMT or geographic targeting to ration beneficiaries among households with young children perform well, although not as well as pure PMT. Comparing the inclusion and exclusion error results with the poverty results shows that the outcomes depend on the metric used. Many fewer poor are covered by the mixed method approaches compared with PMT, but the impact on poverty is not as dissimilar.

Figure 5.10   Reduction in the Poverty Headcount Ratio and Poverty Gap
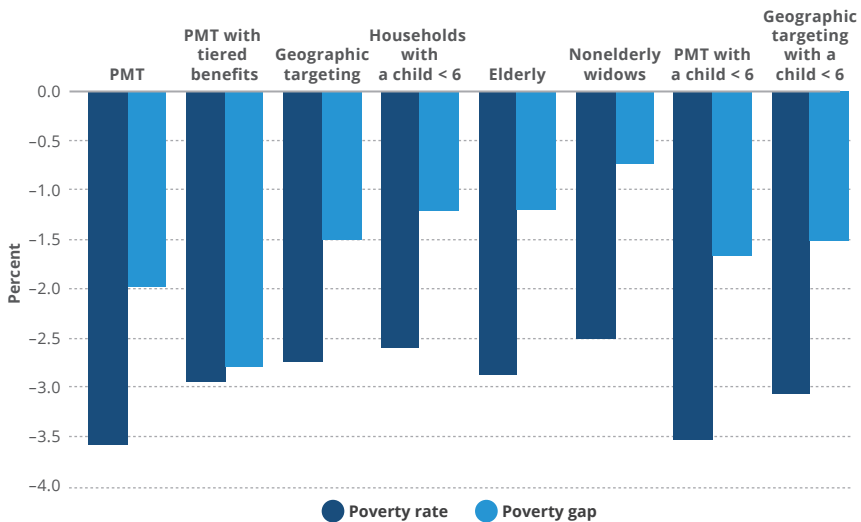
Source: Based on data from the Global Monitoring Database.
Note: PMT = proxy means testing.

The simulation results provide key inputs to policy makers on the trade-offs between different eligibility methods. The analysis confirms the global results from the earlier demographic analysis presented in figure 5.4; in no country would categorical targeting of young children result in an inclusion error lower than 50 percent, and here it would be 62 percent. However, the results suggest that the different methods are close enough to consider the trade-offs between accuracy and implementation. PMT is projected to achieve the largest reductions in the poverty rate and the poverty gap. PMT with tiered benefits improves the lot of the poorest (gap) but pulls fewer people over the poverty line (rate). However, prioritizing children younger than age six years and rationing benefits with a PMT achieves similar results, while prioritizing children younger than age six years in the poorest areas is not too far behind. Pure categorical targeting based on geography or on the combination of geography and demography (assigning benefits only for households with children younger than age six) also both reduce poverty. The analyst has now provided the policy maker information to weigh different considerations. PMT requires a certain technical and financial capacity to implement, yet it is projected to reduce poverty the most. Focusing on children younger than age six is operationally easier and politically popular, but it does less to reduce poverty. Identifying the poorest areas and providing universal benefits also reduces poverty considerably and is easy to implement, but it can create social tensions when all the people in one area receive benefits while no one in a neighboring area does. Simulations like this do not definitely answer the question of how best to determine program eligibility, but they are an important input into the decision-making process.

Simulations provide ex-ante indications of how different targeting methods (and program coverages and benefits) might reduce poverty; the same simulations can result in very different outcomes in different countries. Many things affect how well poverty reduction programs will work in a given country: the shape of the income or consumption distribution, where the poverty line is, how many people a program covers, and how generous its benefits are. The results from one country cannot be safely extrapolated to another; it is necessary to conduct such analysis afresh to be sure of the magnitudes of the trade-offs. The studies discussed in box 5.8 show the variation in the effectiveness of different methods not only within countries, but also across them.

The findings from the simulations can benefit from a reality check with evidence from implemented programs among suitable comparators. A recurrent theme of large, cross-country benchmarking exercises, as done in chapter 2, or with delineation between the choice of methods, as in Devereux et al. (2017) and Coady, Grosh, and Hoddinott (2004), is variability. Nonetheless, looking at how the simulations compare with

BOX 5.8

Illustrative Lessons from the Simulations Literature

The comparative simulations literature reinforces the usefulness of country-specific work, showing some of the methodological choices and how the range of results can vary between countries. This box refers to two well-known papers: Brown, Ravallion, and van de Walle (2016) and Acosta, Leite, and Rigolini (2011).

Brown, Ravallion, and van de Walle (2016) focus on nine African countries—looking mostly at "hard case" countries with relatively high poverty, small social protection programs, and low social protection administrative capacity at the time (Burkina Faso, Ethiopia, Ghana, Malawi, Mali, Niger, Nigeria, Tanzania, and Uganda). The paper simulates the performance of various proxy means testing (PMT) methods and compares those scenarios with a basic income scheme or transfers using various demographic criteria. The simulations are budget neutral, with a budget sufficient to eliminate the poverty gap. For each scenario, the budget is equally divided by the total number of individuals who resided in designated eligible households and distributed according to their size. Two relative poverty lines (bottom 20 and bottom 40 percent of the population) are used, but the main results for poverty impact are emphasized for the bottom 20 percent of the population. The following are among the key findings:

- Although the contours of the simulation were set with a budget to close the poverty gap if perfectly targeted, none of the targeting methods comes close to doing so. This emphasizes the difficulty of the "targeting problem." None of the methods reduces the poverty headcount by more than 5 percentage points (from 20 to 15 percent) nor the poverty gap by half.
- On average across all the countries and formulations, PMT methods would have about twice the impact on the headcount as categorical methods. In the best performing versions of each model, from a baseline of 20 percent poverty, transfers targeted with PMT reduce the headcount to 15.5 percent. The best demographic scenarios reduce the headcount to about 17 percent, and a uniform transfer across all households reduces it to about 17 percent as well (Brown, Ravallion, and van de Walle 2016, table 11).
- The differences are somewhat more marked when considering the poverty gap, moving the poverty gap from an initial 0.05 to 0.03

continued next page

BOX 5.8 *(continued)*

for the better of the PMTs, in contrast to about 0.04 for both the demographically targeted and uniform transfers (table 5.3 in the main text).

- As the authors said, the PMT helps filter out the nonpoor and thus makes higher transfers available to the poor who would receive transfers, but at the cost of missing some of the poor.
- One set of findings pertains to the details of how the PMT is designed—which variables are included and which methods of inference are used (such as probit or ordinary least squares, quantile regressions or poverty-weighted least squares), a topic taken up in chapter 6. The findings indicate that methods matter, although they do not overcome the essential issue of statistical error in PMTs.
- The study is unusual in exploring more sets of demographic criteria than is typically done. Those related to children reduced poverty and the poverty gap somewhat more than those focusing on the elderly, widows, or disabled.

In actual policy, none of the countries covered in the study uses only PMT for targeting their main programs. In none of the countries are the programs fully national, and all of them use geographic targeting to a degree. They all use communities for selection and/or validation, although community-based targeting is all but impossible to model in simulations such as these. Many of the programs use demographic targeting to a degree to ensure that households with children and/or elderly members are prioritized as well. Where PMT is used, it is in combination with all these other methods (for Burkina Faso, Malawi, Niger, and Tanzania, see Beegle, Coudouel, and Monsalve Montiel (2018); for Ethiopia, see World Bank (2015a and 2015b) and Beegle, Coudouel, and Monsalve Montiel (2018); for Ghana, see Agbenyo, Galaa, and Abiiro (2017); for Mali, see Heath, Hidrobo, and Roy (2020) and World Bank (2018a); for Nigeria, see NASSCO (2019); and for Uganda, see Hickey and Bukenya (2021)).

Acosta, Leite, and Rigolini (2011) compare simulations of means testing and demographic targeting for 13 Latin American countries across a large range of incomes, generally with high inequality, substantial social protection programs, and administrative capacity ranging from medium to high. The simulations compare social assistance programs that are fully categorically targeted (benefits given to all children or

BOX 5.8 *(continued)*

elderly individuals) and programs of equal budget that are confined to the poor within those demographic groups, using means testing as the method to identify the poor. The scenarios are set to distribute 0.5 percent of gross domestic product and consider an absolute poverty line of $2.50/per person/day, a threshold that counted about 15 percent of the Latin American and Caribbean population as poor in 2010. The authors add to the simulation potential errors in the means testing (an exclusion error of 30 percent of the poor), since no program is perfectly targeted, and some extra administrative cost that would reduce the budget allocated for transfers in poverty targeted programs.

- Categorical targeting can produce strikingly different results depending on the age group selected and the patterns of poverty. On average, categorical transfers to children are 1.6 times more effective in reducing poverty than categorical transfers of equal budget to the elderly, and poverty-targeted transfers are twice as effective. The reasons are straightforward: poverty rates among the elderly are, on average, lower than for children, and poorer families have more children but not more elderly people. The simulations also suggest that the common belief that cash transfers to the elderly can substantially reduce poverty by trickling down to all family members has limited validity: with fewer elderly than children living in poor households, for the trickle-down effect to be effective, money should be transferred across family members living in different households, which is a much less likely event.
- There is also considerable variation across countries in how the methods compare. In Nicaragua, a perfectly targeted program would reduce poverty rates twice as much a categorical one, while in Colombia (the other extreme) this ratio jumps to 7.1. These differences are not explained by income levels alone: effectiveness in Nicaragua and Argentina, two countries with very different income levels, is very similar. Rather, differences in impact depend on a more complex combination of factors, such as how widespread are pockets of poverty with people far off the poverty line.

Both studies reflect on the extent to which the gains from household-specific targeting are substantial enough to overcome some of the challenges. Brown, Ravallion, and van de Walle (2016) show that even within the set of PMT or demographic methods, there are significant differences in the results that can be expected depending on how the criteria are set. Acosta, Leite, and Rigolini (2011) show how different the results from the same methods can be across countries. Both of these findings imply the importance of doing country-specific analysis.

experience in programs selected to be similar along one or more dimensions—design of the program; country context in terms of poverty, inequality, and maybe demographics; and/or implementation capacity and approach—can be helpful.

A small vein of experimental research seeks to compare household-specific methods using comparative treatment arm designs with PMT, HEA, and/or CBT and self-targeting, which is particularly valuable since CBT and self-targeting are the methods that are the most difficult to simulate. Work by Alatas et al. (2012, 2016) in Indonesia; Premand and Schnizter (2018) and Schnitzer (2019) in Niger; Pop in Ghana (2015); Stoeffler, Mills, and del Ninno (2016) for Cameroon; Sabates-Wheeler, Hurrell, and Devereux (2014) for Kenya; and Escot (2018) and Kameli et al. (2018) for Mali are the classics. Dervisevic et al.'s (2020) study of the Lao People's Democratic Republic is not exactly a comparative treatment arm design, but it sheds similar light on the issue. Some of the findings are as follows:

- In Indonesia, PMT is found to be slightly more accurate than CBT as defined against poverty ($2/day). Self-targeting in the application process led to richer households selecting out while those poorer households who also selected out were more likely to be excluded by PMT (meaning self-targeting did not increase exclusion errors beyond those already driven by PMT).
- In Northern Cameroon, PMT is found to be more accurate than CBT as defined against per capita consumption and alternative measures (household food insecurity, multidimensional poverty, and community perceptions) and thresholds.
- In Niger, PMT can more effectively identify households suffering from persistent poverty, but HEA is better for identifying those suffering transient food insecurity. PMT and CBT performed similarly on food security, asset ownership, income per capita, and malnutrition.
- In Ghana, in a very small study, PMT was slightly more accurate than CBT.
- In Lao PDR, village heads performed about the same as a PMT in selecting poor women to participate in a public works program.
- In Kenya's Hunger Safety Net Program[23] phase 1 assessment, PMT was predicted to perform better than CBT at identifying the poorest and food insecure households. CBT performed better than indicator targeting for a social pension (individuals ages 55 years and older) and a dependency ratio (households with a dependency ratio above a certain threshold).
- In the region of Gao in Mali, HEA[24] did not seem to distinguish the type of vulnerability,[25] although it was meant to select only food insecure households. The authors also found that the results improved when

using a PMT[26] on an independent data collection and not over the pre-list of households that were preidentified by humanitarian agencies operating in the Gao region.

- Communities may bring legitimacy to the process in some places but not universally. In Indonesia, communities showed greater acceptance of CBT than PMT. In contrast, in Niger, PMT was shown to have a slight preference due to past experiences of community bias in CBT.
- Communities appeared to use definitions of poverty that were not strictly related to household consumption. In Indonesia, the definition appeared to be more related to earnings potential and similarly in Cameroon. In Ghana, communities favored smaller households with fewer prime-aged, able-bodied persons, while the PMT favored larger poor families with two or more working adults.
- In Niger, both CBT and PMT aligned equally (and moderately) well with self-perceptions of poverty. In Indonesia, CBT aligned better than PMT with self-perceptions of poverty.

Although PMT comes off slightly preferred in several of the experiments, in most cases, the differences in accuracy are relatively small. The differences are much smaller than the kinds of exclusion that result from program budgets that are much smaller than the population needs. Moreover, the results depend somewhat on which metric of welfare is used and on the criteria for judgment—targeting accuracy versus community acceptability.

Summary

The literature is not definitive on the choice of methods and moreover, the context always matters. Context includes more technical factors such as the goals of the program, the shape of poverty and inequality, the degree of formality, and administrative capacity. It also includes the less tangible institutional history and political economy. No single method dominates across contexts and evaluation criteria. Table 5.4 summarizes the main methods, when each is appropriate, the minimum conditions for using them, their pros and cons, and how useful they are during shocks. However, the appropriate mix of methods or selection for each country depends on specific historic and political factors, and thus decisions on targeting methods remain a source of discussion in social assistance policy.

The combination of contextual factors—administrative capacity, budget, and the strength or form of social contract—is also key. In countries with very low capacity, and especially those with low capacity, low inequality, and social tensions, developing household-specific targeting will be harder

Table 5.4 Appropriateness and Minimum Conditions of Different Targeting Methods

Method	Geographic targeting	Demographic targeting	Means testing	Hybrid means testing	Proxy means testing	Community-based targeting	Self-targeting	Lottery
What is it?	Location determines potential eligibility for benefits When working in isolation, all the population living in the area of intervention are considered eligible However, it is commonly used as a first phase in targeting, to allocate caseload, with some other method used to further reduce the pool of the eligible	Uses age or other demographic characteristic to determine eligibility for benefits Can be applied in isolation or as an additional criterion in mixed methods When used alone, sometimes referred to as universal child grants or social pensions	Compares a measure of the income, consumption, and/or wealth of the social assistance unit (family, household) with eligibility thresholds Often, but not always, verifies a substantial portion of the information with independent sources	Measures and verifies sum of income, consumption, and assets, as in a means testing Imputes the value of other flows where they are not verifiable Imputations often fairly simple—for example, marginal productivity per hectare in agriculture, or unit of livestock, or assumption of a few days a month of low-wage employment for day laborers—although more sophisticated estimations may be used	Uses easy-to-verify characteristics or proxies (for example, composition of the social assistance unit, size, quality or location of its dwelling, its assets) to predict money-metric well-being Weights derived from statistics/econometric models of various degrees of sophistication	Uses organized local-level groups composed by local leaders, civil society, and government officials; group members are from and are very active in the community, and they decide who in the community should benefit	Anyone may participate, but some element of the program makes it unattractive to the less needy For example, a low wage may be offered in exchange for work on community infrastructure or service projects Nutritious but less preferred foods (broken rice, coarse flours, less attractive packaging)	A random process to ration program resources or slots for enrollment among individuals, households, or communities that are all eligible
When is it appropriate?	When differences in poverty, vulnerability, or implementation capacity have a sharp geospatial gradient May work best when people don't move too often or easily between the delineated areas	To address right base approach When the program is focused on biological or social vulnerabilities of children or elderly When age or family structure are highly correlated with poverty or vulnerability to poverty	When income, consumption, or wealth are relatively easy to verify—for example, through data matching with other government-held records When labor markets are highly formal	When a moderately high share of incomes—but not all—can be verified When labor markets have moderate informality	When informality is too high to make means testing or hybrid means testing viable but household-specific rankings are still desired	To address program administrator's myopia and lack of knowledge about the community To promote community engagement and improve accuracy at lowest level	When the program design is conducive	To address the fact that program's budget is not enough for covering the total number of eligible claimants When it is difficult to rank among many similar claimants, or when such rankings would not be socially accepted

Table 5.4 *(continued)*

Method	Geographic targeting	Demographic targeting	Means testing	Hybrid means testing	Proxy means testing	Community-based targeting	Self-targeting	Lottery
Minimum conditions	Data to build geospatial analysis on indicators such as well-being, poverty, social development, access to services, infrastructure, climate, soil, and so forth, including big data	Good coverage of identification documents to verify age For poverty reduction, monetary poverty or vulnerability must be highly correlated with the predefined category	High levels of literacy and documentation that can be used as proof of declared information; high capacity levels of staff to properly collect the information required and to digitize the self-declared information Has effective verification process, including home visits and/or interoperability of government databases	High levels of literacy and documentation that can be used as proof of declared information High capacity levels of staff to properly collect the information required and to digitize the self-declared information Has effective verification process, including home visits and/or interoperability of government databases Plausible models can be built to impute income from household assets or informal labor	Administrative capacity to interview on demand and potentially eligible applicants and/or to conduct survey sweeps in high-poverty areas	Requires a strong, small, and cohesive community structure (hard to use for larger groups where knowledge of one another is limited) Requires effective outreach and capacitation to local actors that will be running the process and supporting program implementation	Subsidies: clear dichotomy in place so that the selected goods are not attractive to the nonintended population but available to the intended population Temporary employment; type of work or benefit amount or goods is not attractive to the nonintended population	Have a transparent process to randomly select the beneficiaries
Pros	Simple to apply and does not create social tensions among close neighbors, though it may across jurisdictions Errors of exclusion are transparent to the political process	Corresponds with notions of deservingness in most places Unlikely to be stigmatizing Relatively simple to implement and for people to understand	Accurate metric for well-being when its development follows basic standards and minimum conditions It is sensitive to quick changes in well-being, either idiosyncratic or covariate	Reliable metric for predicting full well-being when its development follows basic standards and minimum conditions It is somewhat sensitive to quick changes in well-being; the formulae for imputations for informal incomes may need to be adjusted in response to covariate shocks	A statistically plausible and replicable method to rank households when informality is high	Benefits from the locals and their knowledge of the community to identify the population of interest Generates local level buy-in because the locals feel they are part of the process; improves acceptability of the program	Little administrative effort given to eligibility (other aspects of running public works programs are demanding, as are the logistics of food distribution)	Transparency Fast and inexpensive; replicable in any environment, both rural and urban Requires minimal administrative capacity

Table 5.4 *(continued)*

Method	Geographic targeting	Demographic targeting	Means testing	Hybrid means testing	Proxy means testing	Community-based targeting	Self-targeting	Lottery
Cons	There are likely to be clear errors of exclusion as poverty, vulnerability, or other forms of need will exist in excluded territories Can encourage migration to the program's treated areas	May be only mildly correlated with money-metric measures of welfare	High requirements for data available for verification Relies on qualified administrative staff	Relies on quality of data available for verification Relies on qualified administrative staff Contains some statistical error from the imputations	Contains statistical error; formulae may be more or less precise depending on context and statistical methods Difficult for people to understand; lack of transparency Insensitive to quick changes in well-being	Local actors can have own preferences and consequently bias the process toward certain groups Different communities may implement guidelines differently May reinforce existing power structures and replicate patterns of exclusion Social tensions within the population and local actors can arise	May not reduce the number of people desiring to participate enough to meet a budget or implementation capacity; additional layers or filters or other eligibility criteria may be needed	Not to be used when population is heterogeneous enough so that the difference between two eligible people does matter and there are reasonable means to rank more finely
Shock responsiveness	Often used to respond to natural disasters May be used to select areas for resilience-building activities in disaster-prone areas, or in association with early warning system	Not shock-responsive by nature, but if high coverage programs exist, benefits can be increased quickly	It is sensitive to quick changes in well-being, either idiosyncratic or covariate	Moderately sensitive to quick changes in well-being	Insensitive to quick changes in well-being, and consequently not suitable for addressing shocks and for shock response without adaptations	It seems that communities often incorporate local knowledge of which households are suffering idiosyncratic shocks such as illness, accident, or unemployment; for covariate shocks like natural disaster, may involve a new assessment exercise to update information	People may change their calculus about self-targeting as their need increases or decreases—for example, a family that prospers may no longer report for public works, or one that loses its preceding job or faces a downturn may seek additional days of public works	Can be used for some shock-responsive programs by ensuring that those who feel affected would apply and would have equal chance of entering in the program

Source: Original compilation for this publication.

than average and yield smaller returns. In such settings, using combinations of geographic, demographic, and self-targeting or even lotteries may be a sensible choice. In countries with high formality and well-developed income taxes, means testing will come naturally. These decisions need not be uniform withing countries. Cities may have much higher inequality and higher administrative capacity than the remote rural areas of the same country, and different choices would be appropriate in the two contexts.

The question of whether to use simpler methods, such as geographic and demographic targeting, or to develop household-specific methods must be based on the fit for purpose. Using geographic targeting to select only some areas in which to work may fit well with geographically delineated natural disasters but occasion large errors of exclusion for poverty-oriented programs in normal times. Demographic targeting for some purposes is axiomatically a good fit for purpose, and for poverty-related purposes, it will be inexact, although possibly pragmatic. Further, the fit for purpose can vary by program within a given country. For a school lunch program, geographic targeting to poor areas, possibly excluding any categories of schools (private or upper secondary) that serve students who are less poor, may be appropriate, whereas for a last resort income support program, some household-specific targeting may be used. In countries that have a well-developed method of household-specific assessment, multiple programs that use household-specific assessment may use the same means of assessment, but even in these countries, not all programs will use the household assessment; some may use only the categorical methods.

For household-specific targeting, there is a fairly clear order of preference, although sometimes the context narrows the choice set considerably. In some countries, means testing is feasible, and with no inbuilt statistical errors, it is easily adopted as the first choice. This range of countries may be extended by HMT, which may have some errors in the imputations, but the imputations affect only some households and some of their incomes are still lower than in many methods. The range of countries where such methods are applicable is increasing with the secular trend in data availability and may be applicable in still more countries for programs where the eligibility threshold is set quite high. In some countries with high informality, means testing/HMT will not be able to distinguish the poorest but may be able to rule out the wealthiest. However, in many developing countries, the degree of informality implies that means testing or even HMT will not be accurate; therefore, those choices are often deemed to be off the table. In these cases, PMT, CBT, or some combination becomes the common option. In many places, the community has long been a part of targeting processes and although its role may change from full-out decision making to more supportive roles in outreach, data collection, and monitoring, path dependency will maintain a degree of involvement. Conversely, in some settings, the

degree of community cohesion may not allow CBT. This may be true in urban settings where density and mobility (in both residence and where time is spent during each day) are so high that people do not know their neighbors well, or where geographical communities are socially divided by ethnicity or conflict. The Djibouti case study provides an example of geographic targeting combined with CBT or CBT in rural areas and PMT in urban areas. Still, PMT, CBT, and their combination are methods that are on the table and used in a large share of developing countries. Where these are insufficient or undesired, some rationing, such as by self-targeting, geography, or another observable characteristic, or even by lottery, may still be an option.

Whichever method is selected, it is important to do it as well as possible. One part of that involves all the elements of the delivery chain reviewed in chapter 4. These require careful planning, adequate resources, realistic implementation plans, coordination, and a plan for building capacity, learning by doing, and adjusting. They also involve the details of the targeting method itself—the specific data and methods of inference to be used, which is the topic of chapter 6.

Welfare targeting systems often evolve as constraints change and social policy develops. Constraints can change in response to capacities built by social protection programs and as general secular trends in the economy or governance change. Programs and goals can evolve, usually from simpler or quick and dirty to more elaborate or precise. Implementing, learning from constant monitoring and periodic process evaluations, and then adjusting are necessary. Adjustments may improve the implementation or accuracy of the original targeting method, but they may also involve shifting individual programs or a whole social registry from one method to another.

Notes

1. Coady, Grosh, and Hoddinott (2004); Devereux et al. (2017); Slater and Farrington (2009).
2. There are limitations to the coverage and quality of these descriptors of program design, which is why ASPIRE has redesigned the formats and processes it will use in a large updating of administrative data in 2022. This is also why so far ASPIRE has not made much use of the data on program design. But in the internal review process for this book, the reviewers voiced an appetite to see some numbers, with all due caveats: (1) there are no targeting data for about a quarter of the programs, although there does not appear to be any marked bias by region or country income level; (2) the coding is done with a different list of categories and some possible unevenness, as described in the text; and (3) the coding of the qualitative variables was mostly done in the first year of the collection of the expenditure data and often not updated, which is a minor

concern. Moreover, the results are not weighted, so a small program and a large program contribute the same to the analysis. In some ways this is not conceptually problematic—a means tested guaranteed minimum income may have lower coverage than a categorical/demographically targeted child allowance, but both are important parts of the country's social protection system and worthy of observation. In trying to be complete in an expenditure sense, ASPIRE tries to capture as much spending as it can and thus captures not only the flagship programs of the main types in each country, but it also often captures many smaller programs. Thus, the results in the database sometimes feel unfamiliar to those used to reading principally about the flagship programs with big name recognition. Finally, the methods cannot be connected to outcomes. As explained in chapter 2, the household data are processed in groupings not at a program-specific level.

3. The coding in ASPIRE does not allow HMT as an option. Programs that might be labeled HMT in the terminology used in this book were largely coded as means testing in the data.

4. Bance and Schnitzer (2021); Blumenstock (2020); Blumenstock, Cadamuro, and On (2015); Jean et al. (2016).

5. Another initiative is the Pula Advisors, a Kenya-based agricultural insurance scheme that uses rainfall data collected by satellites to estimate the amount of precipitation for relatively small areas to which farmers can be matched. Using machine learning algorithms, Pula aims to provide individual farmers insurance rather than geographical-level area insurance, providing tailor-made protection against adverse growing conditions and thus protecting them more effectively from income shocks. However, Ohlenburg (2020) indicates that Pula still needs detailed data collection at the household level to counteract unrepresentative data, as its modeling was biased toward larger farms that typically have higher and more stable yields due to the limited availability of data for small farmers with variable yields.

6. Center for Effective Global Action at the University of California, Berkeley (https://cega.berkeley.edu/).

7. https://news.berkeley.edu/2020/06/02/satellite-images-phone-data-help-guide-pandemic-aid-in-at-risk-developing-countries.

8. Evidence suggests that CBT seems to be focused on factors other than monetary poverty, such as possession of livestock and land, human and physical capital asset holding, and household earning capacity (Alatas et al. 2012; Karlan and Thuysbaert 2016; Stoeffler, Mills, and del Ninno 2016).

9. See Holzmann et al. (2008) for a description of the HEA.

10. Lotteries are used as well as in other public policies ranging from school admissions (for example, the Federal Pedro II school for basic education and grades 1 and 2 in Brazil or the Prince George's County Public School Specialty Programs in Maryland in the United States) to special visa allocation (for example, the Diversity Visa Program in the United States).

11. The programs can be set up in various ways—giving more or less emphasis to the income support the workers achieve versus the public investment value of the works done, whether there is any attempt to provide training or increased chances of private sector employment, and whether there is a guarantee or

not. (See, for example, Gentilini et al. [2020, chapter 2] or Subbarao et al. [2013] for more on public works programs generally.)

12. https://www.worldbank.org/en/news/feature/2021/03/08/enabling-women -to-work-and-their-children-to-blossom-the-double-success-story-of-mobile -childcare-units-in-burkina-faso.

13. For more on energy subsidy analysis, see Younger, Osei-Assibey, and Oppong (2017) for Ghana; World Bank (2015c) for Indonesia; Martinez-Aguilar (2019) for Mexico; Arunatilake, Jayawardena, and Abayasekara (2019) for Sri Lanka; and Younger (2019) for Tanzania.

14. Occasionally, an economic shock may affect certain sectors of the economy much more than others and these may be spatially concentrated. For example, a fall in international coffee prices will affect the zones of countries where coffee is a major crop more than other parts of the same countries.

15. In Zambia, a HelpAge (2009) report underscores the problems caused by lack of an open registration process for social pensions in the Katete region, and exclusion of anyone who turned 60 (the cutoff for receiving pensions) after the initial registration was completed. An assessment of the Old Age Allowance in Bangladesh found that 24 percent of households receiving the grant did not have a member who was age 60 or older (Slater and Farrington 2009).

16. The Global Monitoring Database is the World Bank's repository of multitopic income and expenditure household surveys, which are used to monitor global poverty and shared prosperity. The household survey data are typically collected by national statistical offices in each country and then compiled, processed, and harmonized. https://www.worldbank.org/en/topic/poverty /brief/global-database-of-shared-prosperity.

17. The focus here is on the income/consumption measures that are most used in poverty targeting, although the methods can be used analogously for other welfare measures, such as an asset index, calorie consumption, and food security/consumption.

18. Pairing means testing and PMT may initially seem a bit paradoxical but is not necessarily. For example, a means test can be done for those with formal income and a PMT for those without observed formal income.

19. The infrequent updating of registry data is often due to the use of time-consuming and costly survey sweeps. Such a practice is not inherent in PMT and there is no reason the registry data cannot be updated more frequently using other approaches. However, the relatively static nature of the underlying variables used in the PMT model makes PMT less responsive to shocks.

20. Evidence from Indonesia suggests that the degree of welfare loss from elite capture is relatively low (Alatas et al. 2019).

21. The Distribution Characteristic Index is also a useful measure in simulations. It not only assesses greater value for benefits received by the very poor compared with the just poor, but also provides some value (although less) for benefits received by those just above the poverty line (unlike the poverty gap). However, it is more complex to explain to policy makers and the public.

22. From chapter 2, the poverty rate is the percentage of people who are below the poverty line, and the poverty gap is how far below the line they are on average.

23. OPM (2011).
24. Escot's (2018) findings indicate that HEA in Gao was the most appropriate method for an emergency response as it managed to be implemented quickly, guaranteeing the acceptance of leaders but potentially at the expense of reliability.
25. Food insecurity, low human and physical capital asset holding, and low household earning capacity.
26. Escot's (2018) findings indicate that PMT, which seems less suited to an emergency, has better reliability and its long-term perspective seems better suited for long-term interventions.

References

Abelson, Brian, Kush R. Varshney, and Joy Sun. 2014. "Targeting Direct Cash Transfers to the Extremely Poor." DataKind DataCorps, New York. http://dx .doi.org/10.1145/2623330.2623335.

Acosta, Pablo, Phillippe Leite, and Jamele Rigolini. 2011. "Should Cash Transfers Be Confined to the Poor? Implications for Poverty and Inequality in Latin America." Policy Research Working Paper 5875, World Bank, Washington, DC. https:// openknowledge-worldbank-org.libproxy-wb.imf.org/handle/10986/3642.

Agbenyo, Fauster, Sylvester Zackaria Galaa, and Gilbert Abotisem Abiiro. 2017. "Challenges of the Targeting Approach to Social Protection: An Assessment of the Ghana Livelihood Empowerment against Poverty Programme in the Wa Municipality of Ghana." *African Journals OnLine* 14 (1). https://doi.org/10.4314 /gjds.v14i1.2.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2016. "Self-Targeting: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 124 (2): 371–427. https://doi .org/10.1086/685299.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2019. "Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia." *AEA Papers and Proceedings* 109: 334–39. https://www.aeaweb.org/articles?id=10.1257/pandp.20191047.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 102 (4): 1206–40. https://www.aeaweb.org /articles?id=10.1257/aer.102.4.1206.

Alderman, Harold, Ugo Gentilini, and Ruslan Yemtsov. 2018. *The 1.5 Billion People Question: Food, Vouchers, or Cash Transfers?* Washington, DC: World Bank.

Arruda, Pedro Lara. 2018. "Malawi's Social Cash Transfer Programme." One Pager 398, International Policy Centre for Inclusive Growth, Brasilia, Brazil. https:// ipcig.org/sites/default/files/pub/en/OP398_The_fragmented_operation_of_ Malawi_s_Social_Cash_Transfer_Programme.pdf.

Arunatilake, Nisha, Priyanka Jayawardena, and Ashani Abayasekara. 2019. "CEQ Master Workbook: Sri Lanka (2009–2010)," CEQ Data Center on Fiscal

Redistribution, CEQ Institute, Tulane University and World Bank, Washington, DC. https://commitmentoequity.org/datacenter/.

Atamanov, Aziz, Jon Jellema, and Umar Serajuddin. 2015. "Energy Subsidies Reform in Jordan: Welfare Implications of Different Scenarios." Policy Research Working Paper 7313, World Bank, Washington, DC. http://hdl.handle.net/10986/22189.

Bance, Paul, and Pascale Schnitzer. 2021. "Can the Luck of the Draw Help Social Safety Nets?" Social Protection and Jobs Policy and Technical Note No. 24, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/35385.

Banerjee, Abhijit, Rema Hanna, Benjamin A. Olken, Elan Satriawan, and Sudarno Sumarto. 2021. "Food vs. Food Stamps: Evidence from an At-Scale Experiment in Indonesia." NBER Working Paper No. 28641, National Bureau of Economic Research, Cambridge, MA. https://www.nber.org/papers/w28641.

Barnes, Sally-Anne, Sally Wright, Pat Irving, and Isabelle Deganis. 2015. "Identification of Latest Trends and Current Developments in Methods to Profile Jobseekers in European Public Employment Services: Final Report." Directorate-General for Employment, Social Affairs and Inclusion, European Employment Policy Observatory, European Commission, Brussels.

Beegle, Kathleen G., Aline Coudouel, and Emma Mercedes Monsalve Montiel. 2018. *Realizing the Full Potential of Social Safety Nets in Africa*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/657581531930611436/Realizing-the-Full-Potential-of-Social-Safety-Nets-in-Africa.

Besley, Timothy, and Stephen Coates. 1992. "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty Alleviation Programs." *American Economic Review* 82 (1): 249–61. http://www.jstor.org/stable/2117613.

Besley, Timothy, and Ravi Kanbur. 1988. "Food Subsidies and Poverty Alleviation." *Economic Journal* 98 (392): 701–19. https://econpapers.repec.org/article/ecjeconjl/v_3a98_3ay_3a1988_3ai_3a392_3ap_3a701-19.htm.

Bickenbach, Jerome, Aleksandra Posarac, Alarcos Cieza, and Nenad Kostanjsek. 2015. "Assessing Disability in Working Age Population: A Paradigm Shift from Impairment and Functional Limitation to the Disability Approach." World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/22353.

Blumenstock, Joshua. 2020. "Machine Learning Can Help Get COVID-19 Aid to Those Who Need It Most." *Nature*, May 14. https://doi.org/10.1038/d41586-020-01393-7, Epub ahead of print, PMID: 32409767.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76. https://doi.org/10.1126/science.aac4420.

Bottan, Nicolas, Bridget Hoffmann, and Diego Vera-Cossio. 2021. "Stepping Up during a Crisis: The Unintended Effects of a Noncontributory Pension Program during the Covid-19 Pandemic." *Journal of Development Economics* 150: 102635. https://www.sciencedirect.com/science/article/pii/S0304387821000146.

Brodmann, Stefanie, Florencia Devoto, and Emanuela Galasso. 2015. "Interlinking Nutrition and Workfare during the First 1000 Days: A New Social Safety Net in Djibouti." Society for Research on Educational Effectiveness, Evanston, IL.

Brown, Caitlin, Martin Ravallion, and Dominique van de Walle. 2016. "A Poor Means Test? Econometric Targeting in Africa." Policy Research Working Paper 7915, World Bank, Washington, DC. https://openknowledge-worldbank-org .libproxy-wb.imf.org/handle/10986/25814.

Clert, Carine, and Quentin Wodon. 2001. "The Targeting of Government Programs in Chile: A Quantitative and Qualitative Assessment." World Bank, Washington, DC. https://hubs.worldbank.org/docs/imagebank/pages /docprofile.aspx?nodeid=6559394.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/14902.

Coady, David, Ian W. H. Parry, Louis Sears, and Baoping Shang. 2015. "How Large Are Global Energy Subsidies?" Working Paper 15/105, International Monetary Fund, Washington, DC. https://www.imf.org/en/Publications/WP/Issues /2016/12/31/How-Large-Are-Global-Energy-Subsidies-42940.

Conning, Jonathan, and Michael Kevane. 2002. "Community-Based Targeting Mechanisms for Social Safety Nets: A Critical Review." *World Development* 30 (3): 375–94. https://www.academia.edu/1642096/Community_Based_Targeting _Mechanisms_for_Social_Safety_Nets_A_Critical_Review.

Curry, February Amelia. 2019. "Policies and Potentials for Women Empowerment in Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA)." Blog post, October 15. https://socialprotection.org/discover/blog/policies -and-potentials-women-empowerment-mahatma-gandhi-national-rural -employment.

Della Guardia, Anne, Milli Lake, and Pascale Schnitzer. Forthcoming. "Poverty Targeting When Everyone Is Poor: The Unintended Consequences of Selective Inclusion in Cash Transfer Programs." World Bank, Washington, DC.

del Ninno, Carlo, and Bradford Mills. 2015. *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*. Africa Development Forum. Washington, DC: World Bank and Agence Française de Développement. http://hdl.handle .net/10986/21369.

Dervisevic, Ervin, Seth Garz, Aneesh Mannava, and Elizaveta Perova. 2020. "In Light of What They Know: How Do Local Leaders Make Targeting Decisions?" Policy Research Working Paper 9465, World Bank, Washington, DC. https:// openknowledge.worldbank.org/handle/10986/34732.

Desiere, Sam, Kristine Langenbucher, and Ludo Struyven. 2019. "Statistical Profiling in Public Employment Services: An International Comparison." OECD Social, Employment and Migration Working Papers No. 224, Organisation for Economic Co-operation and Development, Paris. https://dx.doi.org/10.1787 /b5e5f16e-en.

Devereux, Stephen, Edoardo Masset, Rachel Sabates-Wheeler, Michael Samson, Althea-Maria Rivas, and Dolf te Lintelo. 2017. "The Targeting Effectiveness of Social Transfers." *Journal of Development Effectiveness* 9 (2): 162–211. https://doi .org/10.1080/19439342.2017.1305981.

Devoto, Florencia, Emanuela Galasso, and Stefanie Brodmann. 2017. "Women at Work: Evidence from a Randomized Experiment in Urban Djibouti." In *Three

Essays in Development Economics, edited by Florencia Devoto. Paris: École des Hautes Études en Sciences Sociales; École doctorale d'Économie.

Drèze, Jean, Prankur Gupta, Reetika Khera, and Isabel Pimenta. 2018. "Casting the Net: India's Public Distribution System after the Food Security Act (May 13, 2018)." https://ssrn.com/abstract=3177691 or http://dx.doi.org/10.2139/ssrn.3177691.

Durable Solutions Platform. 2019. "Far from Home: Future Prospects for Syrian Refugees in Iraq." Durable Solutions Platform, Danish Refugee Council, International Rescue Committee, and Norwegian Refugee Council, Copenhagen, Denmark. https://drc.ngo/media/aatg3dkd/far-from-home-future-prospects-for-syrian-refugees-in-iraq_-january2019.pdf.

Durán-Valverde, Fabio, José F. Pacheco-Jiménez, Taneem Muzaffar, and Hazel Elizondo-Barboza. 2020. "Financing Gaps in Social Protection: Global Estimates and Strategies for Developing Countries in Light of the COVID-19 Crisis and Beyond." Working Paper No. 14, International Labour Organization, Geneva. https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---soc_sec/documents/publication/wcms_758705.pdf.

Escot, Fabrice. 2018 "Le ciblage dans les systèmes de protection sociale: étude comparative sur deux méthodes: HEA et une méthode PMT de ciblage pour des programmes de transferts monétaires dans le cercle de Gao, Mali, 2016-17—A Qualitative Study." http://www.miselimali.org/fs/ECHO/duwo4-Miseli_ECHO_BM_Etude_comparative_ciblage_Rapport_post_atelier_restitution.pdf.

Flochel, Thomas, and Sudarshan Gooptu. 2017. "The Energy Subsidy Reform Assessment Framework: Guidance for Comprehensive Energy Subsidy Reforms." World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/28863.

Gelb, Alan, and Anit Mukherjee. 2019. "Fuel Subsidy Reform and Green Taxes: Can Digital Technologies Improve State Capacity and Effectiveness?" CGD Policy Paper 149, Center for Global Development, Washington, DC. https://www.cgdev.org/sites/default/files/fuel-subsidy-reformand-green-taxes-can-digital-technologies-improve-state-capacity.pdf.

Gentilini, Ugo, Margaret Grosh, Jamele Rigolini, and Ruslan Yemtsov. 2020. *Exploring Universal Basic Income: A Guide to Navigating Concepts, Evidence, and Practices*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/32677.

Gentilini, Ugo, Sarah Laughton, and Claire O'Brien. 2018. "Human(itarian) Capital? Lessons on Better Connecting Humanitarian Assistance and Social Protection." Social Protection and Jobs Discussion Paper No. 1802, World Bank, Washington, DC. http://hdl.handle.net/10986/31011.

Grosh, Margaret, Carlo del Ninno, Emil Tesliuc, and Azedine Ouerghi. 2008. *For Protection and Promotion: The Design and Implementation of Effective Safety Nets*. Washington, DC: World Bank. https://openknowledge-worldbank-org.libproxy-wb.imf.org/handle/10986/6582.

Guven, Melis U., and Phillippe G. Leite. 2016. "Benefits and Costs of Social Pensions in Sub-Saharan Africa." Social Protection and Labor Discussion Paper No.

1607, World Bank, Washington, DC. https://openknowledge-worldbank-org
.libproxy-wb.imf.org/handle/10986/24945.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of
Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya."
Quarterly Journal of Economics 131 (4): 1973–2042. https://doi.org/10.1093/qje
/qjw025.

Heath, Rachel, Melissa Hidrobo, and Shalini Roy. 2020. "Cash Transfers, Polygamy,
and Intimate Partner Violence: Experimental Evidence from Mali." *Journal of
Development Economics* 143: 102410. https://doi.org/10.1016/j.jdeveco.2019
.102410

HelpAge. 2006. "Why Social Pensions Are Needed Now: Briefing on Social
Pensions." HelpAge International, London. https://www.helpage.org/silo/files
/why-social-pensions-are-needed-now.pdf.

HelpAge. 2009. "A Social Pension in Zambia: Perceptions of the Cash Transfer Pilot
in Katete." HelpAge International, London. https://www.helpage.org/download
/4ce3e583ecac0.

Hickey, Sam, and Badru Bukenya. 2021. "The Politics of Promoting Social Cash
Transfers in Uganda: The Potential and Pitfalls of 'Thinking and Working
Politically.'" *Development Policy Review* 39 (S1): 1–20. https://doi.org/10.1111
/dpr.12474.

Holmemo, Camilla, Pablo A. Acosta, Tina George, Robert J. Palacios, Juul Pinxten,
Shonali Sen, and Sailesh Tiwari. 2020. "Investing in People: Social Protection
for Indonesia's 2045 Vision." World Bank, Washington, DC. http://hdl.handle
.net/10986/33767.

Holzmann, Penny, Tanya Boudreau, Julius Holt, Mark Lawrence, and Michael
O'Donnell. 2008. *The Household Economy Approach: A Guide for Programme
Planners and Policy-Makers*. London: Save the Children. https://resourcecentre
.savethechildren.net/node/13676/pdf/hea_guide.pdf.

ILO (International Labour Organization). 2019. *100 Years of Social Protection: The Road
to Universal Social Protection Systems and Floors, Volume I: 50 Country Cases*. Geneva:
ILO. https://www.ilo.org/secsoc/information-resources/publications-and-tools
/books-and-reports/WCMS_669790/lang--en/index.htm.

IMF (International Monetary Fund). 2008. *Food and Fuel Prices—Recent Developments,
Macroeconomic Impact, and Policy Responses*. Washington, DC: IMF. https://www
.imf.org/external/np/pp/eng/2008/090808a.pdf.

Indonesia Ministry of Finance and World Bank. 2015. "Taxes and Public Spending
in Indonesia: Who Pays and Who Benefits?" Working Paper 101667, World
Bank, Jakarta. https://openknowledge.worldbank.org/handle/10986/23600.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and
Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning
to Predict Poverty." *Science* 353 (6301): 790–94. https://www.science.org
/doi/10.1126/science.aaf7894.

Jones, Nicola, Rosana Vargas, and Eliana Villar. 2008. "Cash Transfers to Tackle
Childhood Poverty and Vulnerability: An Analysis of Peru's Juntos Programme."
Environment and Urbanization 20 (1): 255–73. https://doi.org/10.1177/0956247
808089162.

Kakwani, Nanak, and Kalanidhi Subbarao. 2005. *Ageing and Poverty in Africa and the Role of Social Pensions*. Washington, DC: World Bank. https://openknowledge .worldbank.org/handle/10986/8535.

Kameli, Yves, Ahmadou Sangho, Pascaline Rollet, Joseph Meunier, Kassoum Kone, Boureima Belem, Sonia Fortin, and Mathilde Savy. 2018. "Étude comparative des méthodes de ciblage utilisées pour identifier les ménages les plus pauvres dans le cercle de Gao, Mali." Institut de Recherche pour le Développement, Marseille, France. https://documents1.worldbank.org/curated/en/986441 588933634885/pdf/Etude-Comparative-des-Methodes-de-Ciblage-Utilisees -pour-Identifier-les-Menages-les-Plus-Pauvres-dans-le-Cercle-de-Gao-Mali.pdf.

Karlan, Dean, and Bram Thuysbaert. 2016. "Targeting Ultra-Poor Households in Honduras and Peru." *World Bank Economic Review* 33: 63–94. http://hdl.handle .net/10986/34289.

Kidd, Sarina. 2016. "If You Have Only Dust in Your Hands, Then Friends Are Far: When They Are Full, They Come Closer: An Examination of the Impacts of Zambia's Katete Universal Pension." Issue 22, Pathways' Perspectives on Social Policy in International Development, Development Pathways, London. https:// www.developmentpathways.co.uk/wp-content/uploads/2016/07/Katete -Impacts-PP22-1.pdf.

Kim, Lydia, and Ikuko Uochi. 2021. "Distributional Impacts of the COVID-19 Pandemic in Mongolia." Draft Report, East Asia and Pacific Poverty and Equity Practice, World Bank, Washington, DC.

Kojima, Masami. 2013. "Petroleum Product Pricing and Complementary Policies: Experience of 65 Developing Countries Since 2009." Policy Research Working Paper 6396, World Bank, Washington, DC. http://hdl.handle.net/10986/13201.

Kojima, Masami. 2021. "Subsidizing Bottled Gas: Approaches and Effects on Household Use." World Bank, Washington, DC.

Larrañaga, Osvaldo. 2005. "Focalización De Programas En Chile: El Sistema CAS." Social Protection Discussion Paper 0528, World Bank, Washington, DC. ttps:// documents1.worldbank.org/curated/en/664351468023113124/pdf /327580revS.pdf.

Leite, Phillippe, Tina George, Changqing Sun, Theresa Jones, and Kathy Lindert. 2017. "Social Registries for Social Assistance and Beyond: A Guidance Note and Assessment Tool." Social Protection and Labor Discussion Paper No. 1704, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle /10986/28284.

Letswalo, Ofentse. 2020. "Youth and Persons with Disabilities to Be Placed at the Core of EPWP Projects." *Expanded Public Work Program Newsletter*, March, Department of Public Works and Infrastructure, Government of South Africa. http://www.epwp.gov.za/news_room/Newsletter/2020/EPWP_Newsletter _March2020.pdf.

Lindert, Kathy A., Colin Andrews, Chipo Msowoya, Boban Varghese Paul, Elijah Chirwa, and Anita Mittal. 2018. "Rapid Social Registry Assessment: Malawi's Unified Beneficiary Registry (UBR)." Social Protection and Jobs Working Paper 1803, World Bank, Washington, DC. http://hdl.handle.net/10986/31012.

Loxha, Artan, and Matteo Morgandi. 2014. "Profiling the Unemployed: A Review of OECD Experiences and Implications for Emerging Economies." Social Protection

and Labor Discussion Paper No. SP 1424, World Bank Group, Washington, DC. https://openknowledge.worldbank.org/handle/10986/20382.

Machado, Anna Carolina, Charlotte Bilo, Fábio Veras Soares, and Rafael Guerreiro Osorio. 2018. *Overview of Non-Contributory Social Protection Programmes in the Middle East and North Africa (MENA) Region through a Child and Equity Lens.* Brasília and Amman: International Policy Centre for Inclusive Growth and UNICEF Middle East and North Africa Regional Office. https://ipcig.org/sites/default/files/pub/en/JP18_Overview_of_Non_contributory_Social_Protection_Programmers_in_MENA.pdf.

Martinez-Aguilar, Sandra. 2019. "CEQ Master Workbook: Panama (2016)." CEQ Data Center on Fiscal Redistribution, CEQ Institute, Tulane University, New Orleans, LA. https://commitmentoequity.org/datacenter/.

McCord, Anna. 2013. "Community-Based Targeting in the Social Protection Sector." Working Paper 514, Overseas Development Institute, London. https://odi.org/en/publications/community-based-targeting-cbt-in-the-social-protection-sector/.

Mendiratta, Vibhuti, Amr Moubarak, Gabriel Lara Ibarra, John van Dyck, and Marco Santacroce. 2020. "Expansion of Djibouti's National Family Solidarity Program: Understanding Targeting Performance of the Updated Proxy-Means Test Formula." Social Protection and Jobs Discussion Paper No. 2006, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34447.

Merttens, Fred, Jayakrishna Upadhyay, Nupur Kukrety, Dr. Shrochis Karki, and Zara Majeed. 2017. "Lessons from the UNICEF Nepal Emergency Cash Transfer Programme through Social Assistance, Evaluation Findings on the Experience and Learnings of the ECTP." Oxford Policy Management and United Nations Children's Fund, Oxford, UK, https://www.unicef.org/nepal/reports/lessons-unicef-nepal-emergency-cash-transfer-programme-through-social-assistance.

Mitra, Sophie, Michael Palmer, Hoolda Kim, Daniel Mont, and Nora Groce. 2017. "Extra Costs of Living with a Disability: A Review and Agenda for Research." *Disability and Health Journal* 10 (4): 475–84. https://doi.org/10.1016/j.dhjo.2017.04.007.

Mitra, Sophie, Jaclyn Yap, Justine Herve, and Wei Chen. 2021. "Inclusive Statistics: Human Development and Disability Indicators in Low- and Middle-Income Countries." Policy Research Working Paper 9626, World Bank, Washington, DC. http://hdl.handle.net/10986/35448.

NASSCO (National Social Safety Net Coordinating Office). 2019. "Advancing Social Protection in a Dynamic Nigeria." NASSCO, Abuja, Nigeria. http://nassp.gov.ng/wp-content/uploads/2020/04/Advancing-Social-Protection-in-a-Dynamic-Nigeria.pdf.

Ohlenburg, Tim. 2020. *AI in Social Protection—Exploring Opportunities and Mitigating Risks.* Bonn: GIZ. https://www.adb.org/publications/ai-social-protection-exploring-opportunities-mitigating-risks.

Ortiz, Isabel, Matthew Cummins, and Kalaivani Karunanethy. 2017. "Fiscal Space for Social Protection Options to Expand Social Investments in 187 Countries." Extension of Social Security Series Working Paper No. 48, International Labour Office, Geneva, and UNICEF and UN Women, New York. https://www.social-protection.org/gimi/RessourcePDF.action?id=51537.

Packard, Truman G., Ugo Gentilini, Margaret Ellen Grosh, Philip B. O'Keefe, Robert J. Palacios, David A. Robalino, and Indhira Vanessa Santos. 2019. *Protecting All: Risk Sharing for a Diverse and Diversifying World of Work*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en /997741568048792164/Protecting-All-Risk-Sharing-for-a-Diverse-and -Diversifying-World-of-Work.

Paes-Sousa, Romulo, Ferdinando Regalia, and Marco Stampini. 2013. "Conditions for Success in Implementing CCT Programs: Lessons for Asia from Latin America and the Caribbean." Policy Brief No. IDB-PB-192, Social Protection and Health Division, Inter-American Development Bank, Washington, DC. https://publications .iadb.org/en/publication/16347/conditions-success-implementing-cct -programs-lessons-asia-latin-america-and.

Pinstrup-Andersen, Per, ed. 1988. *Food Subsidies in Developing Countries: Costs, Benefits, and Policy Options*. Baltimore, MD: Published for the International Food Policy Research Institute by Johns Hopkins University Press. http://ebrary.ifpri.org /cdm/ref/collection/p15738coll2/id/129517.

Pop, Lucian. 2015. "Options for Improving the Targeting of Safety Nets in Ghana," In *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*, edited by Carlo del Ninno and Bradford Mills. Washington, DC: World Bank. http://hdl.handle.net/10986/21369.

Premand, Patrick, and Pascale Schnitzer. 2018. "Efficiency, Legitimacy and Impacts of Targeting Methods: Evidence from an Experiment in Niger." Policy Research Working Paper 8412, World Bank, Washington, DC. http://hdl.handle.net /10986/29714.

Premand, Patrick, and Quentin Stoeffler. 2020. "Do Cash Transfers Foster Resilience? Evidence from Rural Niger." Policy Research Working Paper 9473, World Bank, Washington, DC. http://hdl.handle.net/10986/34774.

Sabates-Wheeler, Rachel, Alex Hurrell, and Stephen Devereux. 2014. "Targeting Social Transfer Programmes: Comparing Design And Implementation Errors across Alternative Mechanisms." WIDER Working Paper 2014/040, United Nations University–World Institute for Development Economic Research, Helsinki. https://www.wider.unu.edu/sites/default/files/wp2014-040.pdf.

Salehi-Isfahani, Djavad, and Mohammad H. Mostafavi-Dehzooei. 2018. "Cash Transfers and Labor Supply: Evidence from a Large-Scale Program in Iran." *Journal of Development Economics* 135 (C): 349–67. https://www.sciencedirect .com/science/article/pii/S0304387818306084.

Schnitzer, Pascale. 2019. "How to Target Households in Adaptive Social Protection Systems? Evidence from Humanitarian and Development Approaches in Niger." *Journal of Development Studies* 55 (sup1): 75–90. https://www.tandfonline .com/doi/full/10.1080/00220388.2019.1687877.

Silva, Joana, Victoria Levin, and Matteo Morgandi. 2012. *The Way Forward for Social Safety Nets in the Middle East and North Africa*. Washington, DC: World Bank. http://documents1.worldbank.org/curated/zh/462881468052823731/pdf /NonAsciiFileName0.pdf.

Silva, Veronica, Heidi Berner, Luis Díaz Silva, and Javiera Troncoso. 2018. "Registro Social de Hogares de Chile." http://www.desarrollosocialyfamilia.gob.cl/storage /docs/RSH_paper_2.pdf.

Silva-Leander, Sebastian, and Fred Merttens. 2016. "Assessment of Programme Targeting." Hunger Safety Net Programme, Oxford Policy Management, https://www.opml.co.uk/files/Publications/a0013-evaluation-kenya-hunger-safety-net-programme/assessment-of-hsnp2-targeting.pdf?noredirect=1.

Skoufias, Emmanuel, Eric Strobl, and Thomas Tveit. 2017. "Natural Disaster Damage Indices Based on Remotely Sensed Data: An Application to Indonesia." Policy Research Working Paper 8188, World Bank, Washington, DC. http://hdl.handle.net/10986/28365.

Slater, Rachel, and John Farrington. 2009. "Targeting of Social Transfers: A Review for DFID." Overseas Development Institute, London. https://www.researchgate.net/publication/265529981_Targeting_of_Social_Transfers_a_review_for_DFID.

Stoeffler, Quentin, Bradford Mills, and Carlo del Ninno. 2016. "Reaching the Poor: Cash Transfer Program Targeting in Cameroon." *World Development* 83: 244–63.

Subbarao, Kalanidhi, Carlo del Ninno, Colin Andrews, and Claudia P. Rodriguez Alas. 2013. *Public Works as a Safety Net: Design, Evidence, and Implementation*. Directions in Development Human Development. Washington, DC: World Bank. http://hdl.handle.net/10986/11882.

Tesliuc, Emil, Lucian Pop, Margaret Grosh, and Ruslan Yemtsov. 2014. *Income Support for the Poorest: A Review of Experience in Eastern Europe and Central Asia*. Directions in Development Human Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/18886.

Tran, Ahn, Sarina Kidd, and Katherine Dean. 2019. "'I Feel More Loved': Autonomy, Self-Worth and Kenya's Universal Pension." Development Pathways, London. https://www.developmentpathways.co.uk/publications/zambia-katete-universal-pension/.

Tuck, Laura, and Kathy Lindert. 1996. "From Universal Food Subsidies to a Self–Targeted Program: A Case Study in Tunisian Reform." Discussion Paper 351, World Bank, Washington, DC. https://ideas.repec.org/p/fth/wobadi/351.html.

UNICEF (United Nations Children's Fund). 2019. "Realizing Children's Right to Social Protection in Middle East and North Africa: A Compendium of UNICEF's Contributions." UNICEF, New York. https://www.unicef.org/mena/reports/realizing-childrens-right-social-protection-middle-east-and-north-africa.

UNICEF–ODI (United Nations Children's Fund–Overseas Development Institute). 2020. "Universal Child Benefits: Policy Issues and Options." UNICEF–ODI, New York. https://www.unicef.org/reports/universal-child-benefits-2020.

United Nations. 2006. "Convention on the Rights of Persons with Disabilities." Treaty Series 2515, United Nations, New York. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-1-purpose.html.

United Nations. 2015. "Report of the Special Rapporteur on the Rights of Persons with Disabilities." A/70/297. United Nations, New York. https://undocs.org/en/A/70/297.

van Landeghem, Bert, Sam Desiere, and Ludo Struyven. 2021. "Statistical Profiling of Unemployed Jobseekers." IZA World of Labor, IZA Institute of Labour Economics, Bonn, Germany. https://doi.org/10.15185/izawol.483.

Vikan, Selvi, and Katharina Diekmann. 2017. "Leave No One Behind: Mainstreaming Disabilities in National Public Works Programmes." *International Journal for Rural Development*. https://www.rural21.com/english/news/detail/article/leave-no-one-behind-mainstreaming-disabilities-in-national-public-works-programmes.html?no_cache=1.

WHO (World Health Organization). 2001. *The International Classification of Functioning, Disability, and Health*. Geneva: WHO. https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health.

WHO and World Bank (World Health Organization and World Bank). 2011. *World Report on Disability 2011*. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/44575.

World Bank. 2010a. "Egypt, Arab Republic of, Food Subsidies: Benefit Incidence and Leakages." World Bank, Washington, DC. https://imagebank2.worldbank.org/search/12892951.

World Bank. 2010b. *Targeting Poor and Vulnerable Households in Indonesia*. Washington, DC: World Bank. http://hdl.handle.net/10986/26700.

World Bank. 2013. *Djibouti—Crisis Response Employment and Human Capital Social Safety Net Project: Restructuring*. Washington, DC: World Bank. https://imagebank2.worldbank.org/search/18190617.

World Bank. 2015a. *Ethiopia Poverty Assessment 2014*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/21323.

World Bank. 2015b. *Ethiopia—Urban Productive Safety Net Project*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/216981467304914217/Ethiopia-Urban-Productive-Safety-Net-Project.

World Bank. 2015c. "Indonesia Ministry of Finance; World Bank. 2015. Taxes and Public Spending in Indonesia: Who Pays and Who Benefits?" World Bank, Jakarta. https://openknowledge.worldbank.org/handle/10986/23600.

World Bank. 2018a. *Mali—Emergency Safety Nets Project (Jigisemejiri)*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/930561531020867594/Mali-Emergency-Safety-Nets-Project-Jigisemejiri.

World Bank. 2018b. *The State of Social Safety Nets 2018*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/29115.

World Bank. 2018c. *West Bank and Gaza—Gaza Emergency Cash for Work and Self-Employment Support Project*. Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/339391534131064726/West-Bank-and-Gaza-Gaza-Emergency-Cash-for-Work-and-Self-Employment-Support-Project.

World Bank. 2020a. *Disability Policy and Disability Assessment System in Latvia*. Washington, DC: World Bank.

World Bank. 2020b. *Disability Policy and Disability Assessment System in Lithuania*. Washington, DC: World Bank. https://socmin.lrv.lt/uploads/socmin/documents/files/Lithuania_Including%20functioning%20into%20DA_Final_August%202021.pdf.

World Bank. 2020c. *Ethiopia Poverty Assessment: Harnessing Continued Growth for Accelerated Poverty Reduction*. Washington, DC: World Bank. http://hdl.handle.net/10986/33544.

World Bank. 2020d. *Rwanda—First Programmatic Human Capital for Inclusive Growth Development Policy Financing*. Washington, DC: World Bank Group. http://

documents.worldbank.org/curated/en/485501608433311280/Rwanda-First -Programmatic-Human-Capital-for-Inclusive-Growth-Development-Policy -Financing.

World Bank. 2020e. *Strengthen Ethiopia's Adaptive Safety Net Project*. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/439371606618883564 /Ethiopia-Strengthen-Ethiopias-Adaptive-Safety-Net-Project.

World Bank. 2021a. "Additional Financing for Strengthening Ethiopia's Adaptive Safety Net." Project Paper. World Bank, Washington, DC.

World Bank. 2021b. *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank. http://hdl.handle.net/10986/35218.

World Vision. 2015. "Social Cohesion between Syrian Refugees and Urban Host Communities in Lebanon and Jordan." World Vision International, Uxbridge, UK. https://reliefweb.int/sites/reliefweb.int/files/resources/World%20Vision %20International%20DM2020%20Social%20Cohesion%20Report.pdf.

Younger, Stephen. 2019. "CEQ Master Workbook: Tanzania (2011–2012)." CEQ Data Center on Fiscal Redistribution, CEQ Institute, Tulane University, New Orleans, LA. https://commitmentoequity.org/datacenter/.

Younger, Stephen, Eric Osei-Assibey, and Felix Oppong. 2017. "Fiscal Incidence in Ghana." *Review of Development Economics* 21 (4): 47–66. https://doi.org/10.1111 /rode.12299.

Zaidi, Asghar, and Tania Burchardt. 2005. "Comparing Incomes When Needs Differ: Equalization for the Extra Costs of Disability in the U.K." *Review of Income and Wealth* 51 (1): 89–114. https://doi.org/10.1111/j.1475-4991.2005.00146.x.

6

How to Harness the Power of Data and Inference: Technical Discussion for Selected Targeting Methods

Phillippe Leite, Emil Tesliuc, Matthew Wai-Poi, and Margaret Grosh

Data and inference are at the heart of differentiating eligibility or benefits. Such differentiation requires making a judgment about where different people (are likely to) fall along the income spectrum in a data-driven sense. Where optimal data are not available, inference is often used to make approximations. This chapter is about data and inference, how to make choices about them in implementing selected targeting methods. Chapter 5 discusses the different methods and their needs in general terms; this chapter is more detailed and specific and focuses on the how-tos. The chapter might be read iteratively with chapter 5, or prior to reading chapter 5, as the choice of method is best made with a deep understanding of the how-tos of each candidate method.

The technology revolution is changing the landscape of data and its processing, indeed making some things possible today that were not a few years ago and hinting at what may be possible in the not-too-distant future, possible in some places but not others, or possibly for purposes akin to but not quite the same as determining which households should receive social protection benefits or services. It is an exciting time in the world of data

and inference. Thus, the chapter begins with a section on big data writ large and then integrates more details as pertinent in the method-specific sections.

The chapter focuses on the targeting methods that are most intensive in data and/or inference: geographic targeting, means testing, hybrid means testing (HMT), proxy means testing (PMT), machine learning–based extensions of PMT, and community-based targeting (CBT) are discussed, in that order. The chapter gives longer treatment to PMT and its machine learning–based extensions because their heavy reliance on inference requires elaboration on the choices of the algorithms involved. Means testing in contrast, is data intensive, but essentially it relies on addition and subtraction rather than any more complicated math. It may not be possible to do means testing if the available data are inadequate, but it is relatively shorter and simpler to explain the how-tos where data are available. The chapter does not discuss demographic targeting because it requires data—essentially age, maybe civil status, and identification (ID) of some sort—that are not complicated and there is no inference. The sort of basic poverty profiling that is needed to know how age correlates with welfare, if that is desired, is well known and illustrated in chapter 5.

This chapter can be read at multiple levels. It is meant to guide the statisticians, econometricians, or data scientists who may conduct the technical work involved in implementing the detailed design or reform of a targeting method. It is also meant to provide an overview of the choices involved for policy makers whose statistics and mathematics skills are basic or have perhaps grown a bit rusty. These latter readers may not capture every detail in the sections on PMT and machine learning, but they should be able to capture the big ideas.

Some Starting Considerations about Data—Traditional and Big

"Traditional" data for targeting mostly come from household surveys and applicants for programs. The data are gathered expressly to measure welfare. Household surveys such as those used to measure income or consumption and other aspects of the population's welfare, or to provide weights for the consumer price index, can be detailed. However, they are always confined to the sample, periodicity, and questionnaire, which, even when generous by survey standards, are limited relative to the scope and dynamics that are desirable for eligibility determination. Thus, social protection programs that determine eligibility on an individual- or household-specific basis usually mount their own data collection efforts (whether in office or home visits or via virtual channels) as part of the application process. However, because the application data are collected for this sole purpose, it is usual to try to

keep costs low, as part of keeping the administrative costs of eligibility deter-mination reasonable. Economizing on data collection can imply partial mea-surement, infrequent updates, use of proxies, and/or making use of other existing data whenever useful and feasible.

"Big data" are generated to serve other purposes but may also be useful for eligibility assessment and so have the allure of a "free lunch," and their higher frequency of updating means that they may be useful for making eligibility determination much more dynamic. Box 6.1 summarizes differ-ent sources of big data, although the division between those collected by the government and those by commercial firms can be blurred; private

---

BOX 6.1 *(continued)*

Data Generated by Households but Held by Commercial Firms

- *Mobile phone data.* Call detail records that record the frequency of texts and length of calls, as well as the frequency or size of top-ups to data plans.
- *Phone-based location data.* Where people go and when, how long it takes them to get there, and how long they stay there.
- *Social media data.* People's account information on their (self-reported) age, sex, and education; data on the type (and quality) of their device and connectivity; data on their social networks; and data on who they follow, like, or retweet.
- *Commercial financial transactions.* Use of mobile money or credit cards and commercial banking information on income, assets, and debt from mortgage and loan applications (some of this private credit information is reported to the central bank or a public credit bureau, meaning the government may have some access).

---

satellites collect remote sensing data, while in some countries mobile operators may be state owned. Because the data are generated for purposes other than welfare determination, they may be proxies rather than measurements per se, and they may require some inference for use. Nonetheless, many enticing research studies show strong correlations between welfare as mapped by the new sources of big data and more traditional interview-based measures.

An important dimension of big data is who owns them and who may use them for what purpose. Governments using data for eligibility decisions for social protection programs must have access to such data. Ownership is a start, although data privacy laws may prevent or bureaucratic silos may impede sharing data between government agencies or functions. Therefore, a regulatory reform of greater or lesser weight may be needed to make even government-owned data available for use in social protection. Normally a program applicant must give consent as part of the application process for the agency to access and use data from other functions or agencies (social security contributions, property tax registers, and so forth).

A strong regulatory environment is needed to enable reuse of public sector data for eligibility determination. *World Development Report 2021: Data for Better Lives* identifies the elements of such a good practice regulatory environment (World Bank 2021b). These include various legislation governing

open data and access to information, data classification policies, interoperability between government agencies, and licensing arrangements. However, although high-income countries have made significant progress, having adopted around two-thirds of such practices, progress is slower in less developed countries, suggesting that reuse of public sector data to help determine beneficiary eligibility may still take time. Moreover, the content and coverage of such data may be limited relative to its ability to observe welfare at the lower end of the welfare distribution.

A different regulatory or payments problem occurs when using household-generated data owned by private firms. The issues of ownership, privacy, and aggregation across different commercial sources will need to be sorted in ways that are technically feasible, economically practical, and politically acceptable. This is new ground, as yet it is hard to predict how fast or extensively the government will be able to harvest and use such data for social protection purposes. Such developments may be uneven initially when using different sources of data and between countries.

So far, data from e-commerce is not much observed by governments, but because of its parallels with traditional commerce, it might be culturally acceptable to think of building that capacity. If brick and mortar stores pay taxes, the analogy to sellers on electronic platforms is clear, although the details of how to regulate/tax them can be quite complex. Governments also maintain a greater margin of control over private sector data transactions subject to competition or consumer protection laws (World Bank 2021b), which may make them more accessible than other private sector big data. Moreover, governments may in some cases directly observe the transactions if they control the digital currency; for example, the Chinese central bank recently piloted an electronic Chinese yuan (eCNY).[1] Where and when it comes to pass that such data are observable to the government, it is likely that the principal driver will be the tax revenue that the government could generate. If in the process of taxing electronic financial flows, the government converts some part of the data to government-owned data, then the improved observability of welfare may be available for benefit determination as well as taxation. This could be revolutionary in helping to observe the welfare of the informal sector and distinguish welfare across the gradients therein, although where household domestic and business accounts are intertwined, this approach would not completely solve the problem. Observing the purchases by a person who works informally as a day laborer, as domestic staff, or in a larger firm but off the books might yield a good idea of their welfare. But a petty trader may look fairly well-off if the purchases of her stock are taken to measure her welfare; the desired concept would be purchases for private use or the perhaps small profit that the trading yields.

The case for routine government access to household-generated data such as call detail records (CDRs), internet search data, and social media posts is less obvious, other than for selected law enforcement purposes. However, in a growing body of work, some examples of which are provided later in the chapter, such commercial data have been made available to researchers. Those instances show some interesting analyses of patterns of welfare and demonstrate the technical potential to use such data for eligibility determination. Whether it will be deemed culturally acceptable is a discussion that is beginning to unfold, and there may be various practical issues as well.

Some progress has been made in making private sector data available for public purposes through legislation, to support evidence-based policy making and promote innovation and competition (World Bank 2021b). For example, some countries have legislation requiring sharing of private sector data of public interest (OECD 2019), such as from utilities and transportation. A particularly relevant example is France's Law for a Digital Republic, which was enacted in 2016 (French Republic 2016; OECD 2019). This law includes provisions mandating the sharing of private sector data according to open standards for the creation of public interest data sets, which cover data from delegated public services or data that are relevant for targeting welfare payments or constructing national statistics (World Bank 2021b, 214, fn 117).

World Development Report 2021 identifies other approaches to allowing the use of private sector data, including promoting open licensing, data portability, and data partnerships. Open licensing encourages private data owners to invest in mechanisms that provide access to proprietary data in return for control and financial rewards, although this is rare in lower-income countries. Data portability allows individuals to facilitate the transfer of data about themselves between parties. This prevents locking in consumer data and fosters competition between companies. In practice, it means the right to receive a copy of the data from a data collector, the right to transmit the data to another data collector, and the right to request a transfer from one data collector to another. Retaining the same phone number when changing operators is a simple example of data portability. Allowing the transfer of financial data on credit and debit card use, personal loans, and mortgages, as is now done in Australia, is a more complex example. Alternatively, data partnerships are contractual agreements between two businesses or a business and government under a public-private partnership. An example of this is Waze, which provides traffic mobility data to cities and other public organizations for traffic management, emergency response, and other mobility-based projects.[2]

Nonetheless, progress toward the regulatory environment needed to facilitate this voluntary provision of private sector data for public use lags

that of public sector data. The *World Development Report 2021* identifies a regulatory framework for enabling reuse of private sector data, which corresponds to that discussed for public sector data (World Bank 2021b). This includes ID authentication, data portability, and voluntary licensing of access to data. On these dimensions, countries have adopted less than 20 percent of good practices, suggesting that the reuse of private sector data for eligibility determination is further off than that of public sector data.

Moreover, it remains to be seen whether the tech giants who own much of these data would be willing to license them given the likely greater returns to maintaining a data monopoly. Google facilitates 90 percent of all internet searches,[3] which in turn may lead to advertisers paying Google higher prices (Scott Morton and Dinielli 2020). Moreover, much of the value of private sector big data comes from just how big it is. The far greater number of web pages indexed by Google compared with most other search engines means that Google provides better results and thus continues to be popular. The value of social media data comes from the networks they reveal. The greater is the size of these networks, the more the data can reveal and the more valuable they become. Maintaining a near-monopoly on such huge networks of data, such as Facebook does, likely has more value than licensing for other users.

Another important facet of data is the unit of observation. Eligibility for many social protection programs is determined at the household level, although for some programs, it is determined at the individual level, but these are not the usual units of observation for much of big data.

Remote sensing data can be associated with increasingly small geographic areas, but they are still fundamentally about "area" and as such, they are conceptually different from a household or individual. The mapping between the two is sometimes straightforward if people or households have GPS coordinates recorded somewhere or geomappable addresses. But even such links are not perfect. At present, sensing data are often available on a grid that is much larger than a household—a square kilometer is a fine bore. Even where the resolution is smaller, in urban areas, many people can live in a single multistory building. In rural areas, livelihoods can depend on flows from multiple plots in different locations or a mix of farm and off-farm employment. A satellite picture may show that a given field is lush or withered or that the plot is lit or dark at night, but the picture does not show the name of the household head or all the other elements of the household's welfare.

The unit of observation issue also arises in the emerging use of CDR-based PMT to approximate the welfare of households. A problem with this approach is the difficulty of linking phones to individuals and households. First, in many developing countries, poor people often use prepaid subscriber identification module (SIM) cards instead of more costly postpaid

accounts. In many instances, these SIMs are not registered to a particular person (although they are meant to be in theory). Consequently, even if CDR patterns predict a poor person, that person may not be known to the mobile operator or the government. Moreover, often people in developing countries have multiple SIMs from different operators to take advantage of cheaper calls to numbers within different networks or at different times of the day. Usually, these SIMs cannot be matched to the same person, meaning the modeling cannot take their aggregate phone usage into account. Second, many programs are aimed at poor households and not poor individuals. If multiple household members have phones and their usage patterns predict eligibility, the household could end up receiving multiple benefits or at least making the eligibility determination at the household level difficult. Conversely, in particularly poor places, the same phone is used by multiple households, which could confound the models. Over time and with greater training data for machine learning models, it may be possible to resolve some of these issues—for example, household members may be identified by their colocation in the household at night, patterns of communication between each other, and so forth—but the issues raise an additional level of complexity in modeling. They also raise issues of incentives, that is, about how households might alter their use of SIMs or phones.

Using big data requires ways to merge data from one source or database with another. This may require a lot of technical work, but success is increasingly possible, due both to advances in the data ecosystem and increased computing power. A frequent key in data matching between separate administrative records is to use a foundational ID number. As the whole Identification for Development (ID4D) agenda discussed in chapter 4 advances, such easy mergers will become more feasible in the coming years. In the meantime, where foundational IDs are not part of the data sets or have limited coverage among the vulnerable population, data algorithms to match on a combination of keys, such as name, age, sex, address or GPS coordinate, or phone number, and other identifiers can sometimes work well enough. However, the algorithms take significant computing power when the number of individuals to be matched is large and there will still be some failures or mismatches that will require manual processes. An address or even more precise, a GPS coordinate, can merge a household's location with sensing data, although GPS coordinates are usually available only for households where data have been collected in the home in recent years, or where there is a good address system that the government has geomapped.

The complexity of big data has also led to the use of more sophisticated modeling techniques—or machine learning—to understand them. Machine learning algorithms take various forms, which are described later in the chapter. They allow objects to be classified (for example, roof type, paved

roads, and railroads) from satellite imagery or allow the relationship between mobile phone behavior and the phone owner's income to be estimated. Even in the absence of big data, it is important to explore whether the new models may allow more accurate modeling of household income from traditional proxies better than the traditional regression models used in PMT.

Big data and machine learning offer the hope of improving certain aspects of some targeting methods, but they are not a panacea. Big data–driven targeting may sound revolutionary and accurate. Once translated from tech-speak to targeting lingo, it becomes clearer that what is meant is really referring to a poverty map and/or PMT based on different variables or algorithms. This framing makes it is easier to see that the usual questions must be answered. How easy is it to get the data? Do they measure or proxy welfare? How accurately do the formulae predict? How big are the prediction errors? What costs do the methods generate in the usual domains—administrative costs to the program, negative incentives, transaction costs, or stigma for the potential social protection claimant, with respect to political support to the program? How do all these compare with other options? Excitement must be calibrated based on the answers to these questions, which are being actively investigated with new information accruing rapidly but not yet definitively.

Geographic Targeting: Big Data Are Revolutionizing Poverty Mapping

Unlike the other targeting methods discussed in this chapter, geographic targeting does not try to be household or individual specific. Instead, it groups households together at a greater or lesser level of aggregation and supports a treatment differentiated between the resulting groups. It is often used as the first stage of a two-step process, followed by a different method for selecting households or individuals within a selected location.

The poverty map methodology popularized by Elbers, Lanjouw, and Lanjouw (2003) facilitated a wave of geographic targeting. The Elbers, Lanjouw, and Lanjouw (2003) technique was a breakthrough because it found a practical way to combine census data, which can be disaggregated to the lowest geographical level but do not collect information on household income or consumption, with household surveys, which collect income or consumption but are only representative at very gross geographic levels.[4] Econometric models are used to make poverty maps, or small area estimates, for levels of aggregation—district, parish, and municipality—that are much more detailed than the survey's sampling frame (which is often just rural and urban, a few large agroecological zones, or the largest level of administrative unit, such as state). These poverty maps have

been used since the 2000s to help social programs decide where to locate services. For social protection, they are often used to allocate budget or create quotas for the number of people who should be covered by programs in each area.

In recent years, advanced technology for geotagging and processing large databases has allowed researchers and policy makers to explore different uses of small area estimation for supporting social protection and targeting, but mostly these sensing data are used in poverty assessments. The following discussion draws on Areias et al. (forthcoming), which summarizes the application of machine learning to big data to facilitate greater geographical targeting; chapter 8 summarizes the application of machine learning to PMT.

Satellite imagery has been used to produce local area poverty estimation. For example, Engstrom, Hersh, and Newhouse (2021) use a convolutional neural network, which is a machine learning approach, with high-resolution satellite imagery, in Sri Lanka to classify objects (for example, roof type, paved roads, railroads, and so forth) (photo 6.1). They then impute welfare estimates using the Elbers, Lanjouw, and Lanjouw (2003) approach into the 2011 Census of Population and Housing. That is, for each household in the census, per capita consumption is estimated based on models developed from the Household Income and Expenditure Survey using household indicators that are common to both the census and the

Photo 6.1 Example of Classification of Developed Area (Buildings)

a. Raw image   **b. Image with developed area classifier**

Source: Engstrom, Hersh, and Newhouse 2021.

Note: Panels a and b, respectively, show the raw and classified images for developed area classifier from raw satellite imagery. The areas in green show true positive building classifications. The images in red show false positives, areas erroneously classified as buildings.

Household Income and Expenditure Survey. By comparing convolutional neural network classifications with the poverty estimates and predictions of household income, Engstrom, Hersh, and Newhouse (2021) in fact as above explain around 60 percent of the variation in imputed household consumption data averaged at the village level, which could help future geographic targeting without the burden of collecting a census or household-level data. Jean et al. (2016) implement a convolutional neural network with nightlight satellite imagery applied to daytime satellite imagery for Malawi, Nigeria, Rwanda, Tanzania, and Uganda. They find promising results for small area poverty estimates. The authors show that satellite imagery can produce fine-grained poverty and wealth estimates using only available data such as a Living Standards Measurement Study survey, explaining up to 75 percent of the variation in local-level economic outcomes. The approach explains 37–55 percent of the variation in average household consumption (as measured in the Living Standards Measurement Study surveys) and 55–75 percent of the variation in average household asset wealth obtained at the cluster level.[5] In a background paper for the *World Development Report 2021* on data for development, Masaki et al. (2020) show that incorporating satellite data into small area nonmonetary poverty estimates in Sri Lanka and Tanzania improves map accuracy to a degree equivalent to tripling the household survey in Sri Lanka and increasing it by five times in Tanzania.

Mobile phone CDRs have also been used to infer household-level welfare at the regional/cluster level.[6] The prototypical CDR metadata include a hashed phone number of the calling and receiving parties, the type of transaction (call or text), the date and time of the transaction, the cost, the call duration, and an identifier for the cell phone tower used to initiate the transaction, which indicates geographical location. The literature combining CDRs and household surveys provides a promising contribution to predicting low-level poverty and wealth indicators at the cluster level. Blumenstock and Eagle (2010, 2012) in Rwanda and Wesolowski et al. (2012) in Kenya examine correlations between household and individual demographics and call patterns. Blumenstock, Cadamuro, and On (2015) provided the first rigorous machine learning approach to modeling household-level poverty and wealth indicators in Rwanda. Using a phone survey of 856 respondents, they can explain 68 percent of the variation in the first principal component of a principal component analysis wealth index.[7]

Phone data can also be used to create maps of travel times to public services, which account for not just distance but congestion. Roberts, Gil Sander, and Tiwari (2019) show correlations of mobile phone–based traffic congestion time and the times and costs associated with reaching public services such as health and education facilities in Jakarta, Indonesia (see map 6.1). The analysis reveals that regular survey data do not capture the

Map 6.1   Travel Times to Jakarta's Main Government Offices

Travel time
(minutes)

- 1–20
- 21–40
- 41–60
- 61–80
- 81–100
- 101+

PPP $3.10 Poverty rate (%)

0.12   13.57
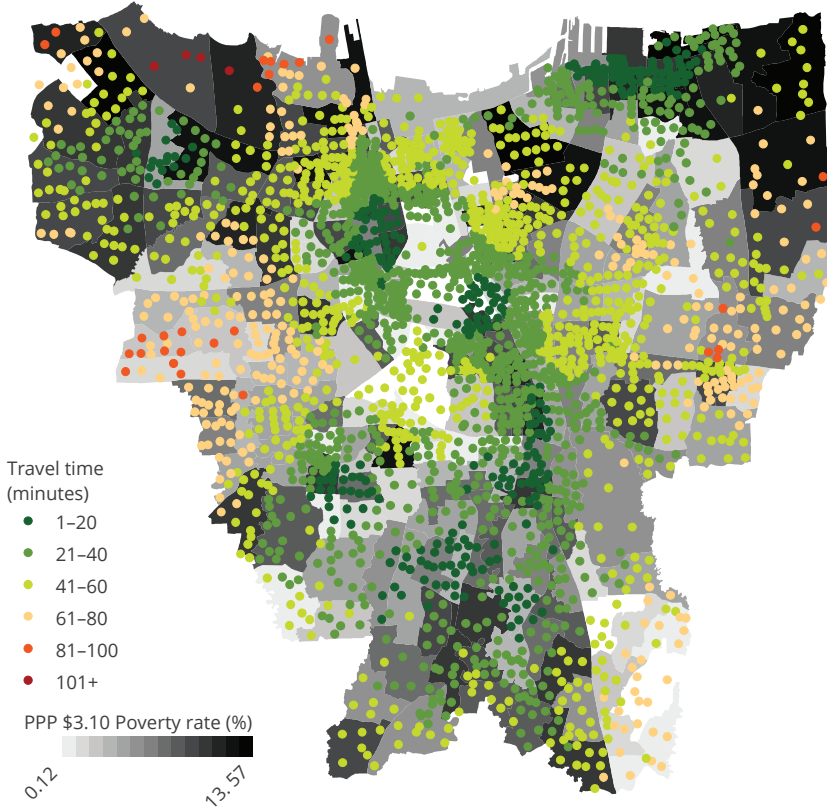
Source: Roberts, Gil Sander, and Tiwari 2019.
Note: The colors on the map correspond to the minimum time to reach a point of interest from the relevant community centroid. PPP = purchasing power parity.

full extent of spatial inequality and constraints on access to services in Indonesia's cities: distance to certain facilities may be an inappropriate indicator of access in settings with high levels of congestion.[8] Recently, CDRs have been used to target individuals directly. This is discussed further in the section on PMT.

However, the use of CDRs faces limitations, particularly in access. In many countries, CDR data are owned by private operators and may not be accessible to governments for use in social protection programs. In countries with multiple operators, access to all firms' data may be necessary to get a nationally representative data set, making access issues even

more difficult. Moreover, as with traditional poverty mapping models, which combine census and household survey data, CDR-based models do well at estimating welfare at the community level but may have much greater errors at the household level, which is the level most pertinent for social assistance eligibility determination.

The use of social media data for poverty assessments is also increasing. Recent work has combined social media data with satellite data to improve or substitute for traditional poverty mapping techniques. For example, Fatehkia et al. (2020) train a model using satellite daytime imagery on Demographic and Health Survey localized (cluster) household principal component analysis wealth indices to predict poverty in India and the Philippines. They then show that use of basic (and freely available) Facebook data can come close to replicating satellite data–based maps or standard poverty mapping approaches or can be used to augment them.[9]

The ever-expanding access to and frequency of big data and the computational power of machine learning provide an encouraging option for prediction that can be used for poverty mapping at fine levels of disaggregation. Chi et al. (2021) have recently combined satellite, CDR, and social media data to construct poverty maps for 135 developing countries. However, there are several significant data constraints. The ability to make inferences for individual households is dependent on the quality of the "ground truth" data from more traditional surveys. Where survey data are unavailable or inaccurate, it will be difficult (or impossible) to develop accurate big data–based models. Hence, investments in big data do not replace the need to collect more information at the household level to improve the ability to differentiate eligibility or benefits. Moreover, the current literature suggests that although identifying droughts and agricultural shocks is feasible with sensing data, identifying economic shocks (such as job loss) and health shocks is not well served by this. Ohlenburg's (2020a) careful analysis of using big data in social protection shows that despite having multiple usages, data protection, including which data the state can legitimately use to determine eligibility (box 6.1), is one among many questions for this promising but still immature field. The Give Directly example discussed in the PMT section offers a cautionary tale about the use of novel data sources. Other researchers, such as Steele et al. (2017), suggest that targeting methods with big data are not yet accurate enough for use at the household level. Nonetheless, in the half-decade since Steele et al. (2017) published that paper, the use of CDRs to target households has moved from the academic to implementation, as discussed later in this chapter.

Moreover, methodological questions remain on data and the accuracy of big data–driven maps. First, several researchers use the Demographic and Health Survey wealth index to train their models (for example,

Blumenstock, Cadamuro, and On 2015; Head et al. 2017; Jean et al. 2016; Masaki et al. 2020). However, this wealth index is the first component from a principal component analysis, which is itself a proxy for the desired measure to be predicted, based on the same household characteristics that PMT uses to predict income or consumption. So, many of the big data models are training on a proxy for the desired measure rather than directly on the measure itself (as models that train on a Living Standards Measurement Study survey or Household Income and Expenditure Survey do). Second, although some big data—notably, administrative and census data—are direct measures, satellite data are modeled and interpolated before they are used in the poverty mapping models. Thus, a combination of noise in the training data and noise in the big data explanatory variables raises some questions about the accuracy of the resulting maps. Moreover, maps are often produced with point estimates and not a measure of precision, which is needed to know how confident a policy maker should be in using the maps at different levels of disaggregation or running the risk of producing biased estimates with inaccurate measures of precision. To date, to our knowledge, the accuracy of the big data approaches is only estimated by simulations in the training data themselves, or in comparison with alternative imputations in the case of Engstrom, Hersch, and Newhouse (2021). There is a need for research that would directly assess the big data maps by, for example, surveying household income or consumption in selected areas with samples that are large enough to compare true local area poverty rates with the predicted rates from the new maps. Future research could also compare the different big data maps—satellite, CDR, and social media—to identify which versions are more accurate.

In addition to determining quotas or regional prioritization for social programs, big data may be used to improve the dynamism of targeting with early warning systems. These data can help policy makers with program response (vertical and horizontal expansion) and planning, especially in times of crises, for better targeted responses. In Bangladesh, for example, in July 2020, the Jumana River experienced more severe and protracted flooding than it had experienced in decades. Data on upstream river levels were used to trigger electronic cash payments to downstream households a few days in advance of the flooding. Subsequent evaluation showed that the anticipatory cash transfer was mostly spent on food and water, and that treated households were 36 percent less likely to go a day without eating during the flood. Three months after the flood, households that had received the transfer reported significantly higher child and adult food consumption and well-being. They also experienced lower asset loss, engaged in less costly borrowing after the flood, and reported higher earnings potential (Pople et al. 2021).

Key Elements for Means Tests

Means-tested programs are typically found in high-income countries with highly formal economies, such as the traditional Organisation for Economic Co-operation and Development (OECD) members. Some countries, notably Australia, Canada, Japan, New Zealand, the United Kingdom, and the United States, employ means tests for most of their social assistance programs. European Union (EU) countries use means-tested programs as a last resort, intended to catch individuals or families whose income is still low after the support provided by the other social protection programs, including generous categorical social assistance programs. Among the income support social assistance programs of 35 European countries, Coady et al. (2021) find that means-tested programs represent about one-third of the programs, or about 1.1 percent of gross domestic product.[10] Within the European Union, southern countries like Greece did not have a means-tested program a decade ago, given the high level of informality. A minimum income program called the Social Solidarity Income, meant to support the poorest households, was rolled out nationally in February 2017 in Greece, bringing the country in line with other EU and most OECD countries.[11]

The landscape of means-tested programs is varied. For example, in the United States, means-tested programs are used extensively to support low-income individuals and families. At the beginning of 2020, before the COVID-19 crisis, they included 79 federal programs that covered about 19 percent of the population.[12] These programs include the Medicaid program, which provides free medical care to low-income adults and children, the elderly and disabled, and long-term care; the Earned Income Tax Credit, which provides a tax credit to families and individuals with relatively low levels of earnings; the Supplemental Security Income program, which provides cash benefits to low-income aged, blind, and disabled individuals; housing subsidy programs, which provide housing vouchers to low-income families, subsidized rent in public housing projects, and support for construction of low-income housing; the Supplemental Nutrition Assistance Program (SNAP), formerly called food stamps, which provides an allotment of funds for food expenditure for low-income families and individuals; the Temporary Assistance to Needy Families (TANF) program, which provides cash assistance for general consumption to low-income families (mostly single mothers and children); school food programs (subsidized breakfasts and lunches for children from low-income families); the Head Start program (providing early education and childcare for children of low-income families); and the Women, Infants, and Children (WIC) program (providing nutritional assistance to mothers, infants, and children at nutritional risk).[13] The extreme fragmentation of the means-tested programs in the

United States reflects the historical development of safety net as well as voter preferences for in-kind programs and concerns about supporting working-age people who might not be working (Moffitt 2015).[14] Other countries, such Australia, have centralized the administration of the programs, and the United Kingdom has consolidated the means-tested programs for working families under a single program, the Universal Credit. Most countries in the European Union have one or few last-resort means-tested program(s); they may use means testing selectively for social pensions, disability support, or labor market programs.[15]

Means-tested programs can be designed for different assistance units: households, families, or individuals. If the assistance unit is the household (or family), total household (family) income is assessed based on the total current income of all members, as well as the incomes that are not attributable to specific members (for example, incomes from agriculture, rental income from jointly owned assets, and social protection family benefits). If the assistance unit is the individual, for example for a means-tested social pension or disability benefit, only the incomes of that individual are considered.

Verified means testing is often considered the gold standard of targeting methods because unlike the other methods, it directly measures the desired welfare concept: the income and sometimes some of the assets of the applicant. Welfare can be measured as the sum of the incomes and assets of the members of the assistance unit. These can be observed and verified by the administration, which has records of such incomes and assets. Assuming full take-up and no underreporting or other types of errors in the administrative databases,[16] a verified means test will have no inclusion or exclusion error. In contrast, HMT, PMT, and geographical targeting cannot observe or verify the full welfare aggregate of the assistance unit—they can only estimate it with some known modeling error. With means testing, there is the potential that by pulling data largely from other existing systems, good targeting accuracy can be achieved.

In practice, means-testing methods rely on measuring a pragmatic subset of incomes and assets rather than trying to quantify every possible source. Thus, the program's definition of income and assets differs from its comprehensive economic definition. Typically, means-tested programs account for formal wages from the main job and sometimes secondary jobs; retirement and other pensions received from pension institutions; regular monthly income from other social protection programs, such as labor market or social assistance programs; and unearned income, such as rents, dividends, court-ordered alimony or child support, and so forth. All these income sources are regularly recorded in administrative or private databases, which are often used to verify the income declared by applicants.

Certain incomes, such as small, irregular, or informal/unrecorded incomes, are typically excluded from the program definition of income. Whether to try to include informal incomes in the program's definition of household income, which represent a small share of total income in the countries where verified means testing is typically used, is more debatable. For example, a low-wage worker may have easily reported and verified wages but pick up money off the books from occasional side gigs like baby-sitting for a neighbor. Not counting such income does undercount welfare. To count it is very difficult, certainly raising transaction costs to participants and administrative costs. Moreover, if such informal income is small, trying to count it may not improve targeting accuracy by much; instead, it could push the low-wage worker into the category of fraudster or discourage work effort, neither of which is desirable. Such sources of income that represent a small proportion of the average total income are irregular or not typically reported; in practice, they are disregarded. Another source that is disregarded, given the difficulty of estimating and verifying it, is consumption from a household's own production. However, if such informal income is larger, more structural changes in method may be needed, such as moving to an HMT or a PMT.

The program administration's ability to observe (the largest proportion of) applicant income and assets implies that it can also observe the change in the level of that income or assets; thus, it can protect those who are vulnerable after a shock. One of the advantages of means testing is that it should be able to detect changes in income and allow households to apply for benefits any time their income falls and receive assistance fairly quickly thereafter. For this to happen, the income sources should be paid and reported at frequent intervals, such as once or twice a month. Thus, a well-off person who becomes unemployed will qualify for assistance from a means-tested program in the next period. This is also straightforward for assets, which are a stock concept (box 6.2). A shock that reduces the stock of a particular asset could be quickly considered, directly (if the program also uses an eligibility line for assets) or indirectly (if the reduction in the stock of a productive asset reduces its income-generating capacity). When used in this way, programs can prevent adjustments from which it is difficult for families to recover—for example, a family losing its housing or selling the car or motorcycle it needs to get the breadwinner to work to make rent—or an outcome with long-term consequences—especially a child raised in hardship long enough to affect the child's growth or development.

Apart from the list of income sources that are included in the definition of income used by the program, the recall period during which incomes should be reported and counted should also be specified. Shorter recall periods, such as a month or a quarter, will allow households that fall on

BOX 6.2

Treatment of Assets in Means-Tested Programs

In many Organisation for Economic Co-operation and Development countries, means-tested social assistance programs use measures of income and assets. Wealth may be measured as the sum of all assets of the assistance unit (the wealth index approach), which is compared to a threshold. Alternatively, the ownership of individual assets, regardless of value, can be used as grounds for declaring a household ineligible (the asset filter approach). Such treatments of wealth are used to exclude asset-rich households from accessing the program regardless of their current income level. To be eligible for a program, households should have both a wealth level below a given threshold and an income below a given threshold.

As with income, this creates disincentives to accumulate and perverse incentives to misreport the level of wealth. To counter these effects, programs may disregard certain assets from the asset test but not others. For example, in Greece,[a] households in the Social Solidarity Income program do not have to sell their houses and vehicles before applying to the program. A household can participate in the program as a homeowner if the total taxable value of the property does not exceed €150,000 (US$170,000), and/or the value of the household's vehicles is estimated at a maximum of €6,000 (US$6,741), calculated according to Article 31 of Law 4172/2013.[b] However, owning other assets, such as a private recreational boat (exceeding 5 meters in length and with engine power exceeding 50 cubic centimeters), aircraft, helicopters, gliders, swimming pools, as well as financial assets over a certain threshold, automatically excludes households from the program, regardless of their income. The same applies to the current value of shares, bonds, and so forth owned over the six months prior to the application, compared with a threshold as a signal of well-being. This implies that households are expected to draw down at least some liquid assets before receiving public support. In the United States,[c] a few states use "broad-based categorical eligibility," which allows low-income Supplemental Nutrition Assistance Program beneficiary households to keep some assets, such as a car if it is used to find and keep a job; a house if the demand to the program is due to a short-term shock (for example, loss of job, divorce, or unexpected temporary disability); and a car and house if the household has seniors or people with disabilities. In Lithuania, households are excluded if the value of their property exceeds the average property value set for the residential

continued next page

BOX 6.2 *(continued)*

area, which is determined by the government according to the norms set for housing and land, as well as other assets including vehicles, livestock, durables, savings, and shares (Tesliuc et al. 2014).

Selecting the right combination of exclusionary asset filters and their thresholds should be done after empirical analysis. Otherwise, the filters could exclude many of the poor in the target group, thus reducing the effectiveness of the program. Programs should use a representative household survey with information on household welfare as well as asset ownership or value. Failing to calibrate the asset filters, without proper ex-ante simulation of their impact, could lead to a high level of exclusion error.

The same consideration applies to programs that use the hybrid means test, which is covered in the next section. Tesliuc et al. (2014) report cases in which setting up such filters relied only on common sense or the beliefs of the social assistance administrators, instead of an empirical assessment, which resulted in large exclusion errors. In Albania's Ndihma Ekonomike, circa 2008, the list of asset filters and conditions was estimated to exclude 90 percent of the poor from the target group in urban areas, especially working-age poor families. This was later corrected; determination of eligibility for the program was switched to a proxy means test in 2018. In Uzbekistan, the imputation coefficients for farming income were set too high and resulted in the exclusion of many poor households. In the Guaranteed Minimum Income program and the Family Allowance program in Romania in 2012, a list of 20 asset filters, which individually excluded no more than 1 to 9 percent of the households in the income target group, collectively excluded one-third of the intended beneficiaries. This situation was subsequently corrected by collecting data for simulations and calibrating the vector of asset filters to minimize the exclusion error.

Thus, the use of asset filters has both pros and cons. The use of exclusionary asset filters may prove effective for improving the focus on extreme or chronic poverty by reducing the inclusion errors and at the same time ensuring the legitimacy of the program. However, if filters are not well set, this approach may lead to exclusion errors that reduce the effectiveness of the program in reducing poverty.

a. World Bank (2019b).

b. Annual objective expense for a passenger car for private use adjusted according to its age: no reduction for a period up to 5 years; a rate of 30 percent for a period of more than 5 years and up to 10 years; and a rate of 50 percent for a period of more than 10 years.

c. Rosenbaum (2019).

hard times to become eligible for public support quickly. This feature will make the program faster in responding to negative income shocks. To maintain a program's adaptability, the recertification or updating of the income of the household should also be frequent. The degree of adaptability will also depend on the frequency of income reporting in the other administrative systems.

There are some practical limits to how quickly information may be updated depending on the concepts and systems. What is practical in a given country would depend on the strength of the program's delivery system (a topic covered in chapter 4), as well as the broader information ecosystem in the country. In general, wages are reported at least monthly or quarterly to the social security agency, but income tax records are filed only annually. In Chile, for example, labor income, pensions, vehicle ownership, and health insurance fees can be verified monthly, while capital income, school fees, and household ownership can be verified annually. In Turkey, the self-reported data included in a citizen's application form are verified both internally and externally. The information in the system is real-time information. It is updated every second. However, when there is an evaluation for the eligibility of social assistance, before each evaluation, the system conducts an online search automatically. This process is completed once a month since social assistance payments are mostly monthly. Nevertheless, the ministry can see the most recent data related to the registered people through active inquiry of the system. Therefore, the means test must be designed considering the many sources of volatility of income and requires significant investments for data management.

By tracking the changes in the income of the assistance unit relatively frequently, means-tested programs can also calibrate the benefit level of the program and operate a large array of benefit formulae. Means-tested programs typically fall into two categories: programs that provide a flat benefit or service, and those whose benefits differ with the income level of the applicant. However, most means-tested programs offer differentiated benefits.

The most frequent means-tested programs that differentiate benefits as well as eligibility are the guaranteed minimum income programs, whose benefits equal the difference between the income level of the applicant and a guaranteed minimum. The guaranteed minimum is often linked to a measure of the subsistence level, which incorporates a range of relevant expenditures and is adjusted to reflect the composition of the assistance unit, per capita or per adult equivalent. For example, most guaranteed minimum income schemes operating in the European Union consider a broad notion of needs, including food, nonfood, and services; in contrast, the US SNAP program—which aims to ensure a minimum level of food consumption for low-income households—aims to fill only the food gap.

Some countries offer one comprehensive benefit, while others use a guaranteed minimum income program as a revolving door to offer additional benefits (for example, heating or housing assistance).

By tracking changes in formal incomes frequently, means-tested programs can operate dynamic benefit formulae, for example, withdrawing benefits as a household's earnings increase. The steepness of the taper is an important design variable. Steeper tapers target more narrowly, and flatter tapers allow more people above the eligibility or poverty line to receive some (reduced) benefits. Steeper tapers (also known as higher marginal tax rates) are expected to reduce work effort more than flatter ones. Such adjustments also have some practical implications for dynamics. In countries where formality is not widespread, individuals in the informal sector may avoid entering the formal sector and losing social benefits when formal labor income is considered. In Brazil, where means testing is used despite informality, government communications campaigns reinforced that the criterion is having an income below the threshold regardless of whether the individual is a formal or informal worker, as many people still believed that formality was an exclusion factor.[17]

Some higher-income countries disregard a certain proportion of the earned income from the program definition of income (they do not include it in the income test) to reduce work disincentives. Beneficiaries whose income hovers around the eligibility line face an implicit 100 percent marginal tax on earnings if benefits are determined as the difference between their income and a guaranteed minimum. For this type of program, a small increase in earnings results in an equal reduction of benefits; hence, the work disincentive may trigger a reduction in the beneficiary's work effort. Partial disregard of the earnings of last-resort income support recipients is allowed by the social assistance legislation in many countries in Europe, such as Cyprus, the Czech Republic, Estonia, Germany, the Netherlands, Portugal, Romania, the Slovak Republic, Slovenia, and Sweden (World Bank 2019a).[18] In Estonia, social assistance beneficiaries' earned income is not considered for the first two months. After two months, 50 percent of earned income is not considered. Portugal disregards a higher share of income (50 percent) for 12 months if a new job was obtained through activation measures. If a new job was found in a different way, 20 percent of the earned income is not considered. Slovenia does not count certain income from informal work, as well as casual and nonrecurring income, when determining eligibility for last-resort income support. Another example is Serbia, where, according to the 2011 Social Welfare Law, for last-resort income support beneficiaries who are able to work, the income they receive from participating in training activities organized by the National Employment Service is not counted as additional income and does not require reassessment of

eligibility for last-resort income support. A large scope of income disregards is found in the Slovak Republic, where 25 percent of wage income, 25 percent of income from occasional work, 25 percent of the activation allowance for voluntary service, and all allowances related to participation in active labor market policies are disregarded in the income test for last-resort income support (MISSOC 2020).

Another parameter of means-tested programs that is used to reduce work disincentives is the gradual phase-out or tapered withdrawal of benefits over time (World Bank 2019a). Beneficiaries can increase their take-home income by combining their earned income with the benefits. Many EU member states allow gradual phase-outs. For example, in Croatia, social assistance beneficiaries receive the full benefit in the first month, 75 percent of the benefit in the second month, and 50 percent of the benefit in the third month. Lithuania continues to pay 50 percent of the previously paid benefit for six months to those graduating from the last-resort income support scheme, to encourage their labor market participation. Hungary modified the design of its social assistance system so that last-resort income support beneficiaries continue to receive some benefits for up to six months after gaining employment. Along similar lines, Latvia's guaranteed minimum income benefit can be received in reduced amounts for a limited duration after securing a salaried job. In Croatia, as of September 2015, last-resort income support beneficiaries who find work can continue receiving benefits in decreasing amounts during the first three months of employment. As of 2013, the Slovak Republic pays part of the last-resort income support benefit together with the wage for 12 months. In France, the minimum income benefit is received when the beneficiary is employed for up to 750 hours per year (for a maximum of 12 months).

Important factors for the success of means testing include the following:

- Databases exist that provide reasonably complete information on the income and assets of the target group and cover the part of the population that is pertinent to the eligibility decision to be made. If the program is meant to be nearly universal and screen out only the top, say, tenth or quarter of the population, this may be possible even where informality is high and data systems have low coverage. Conversely, if the eligibility threshold is set at the bottom 10 percent of the population, a verified means test would require complete information for the entire population.
- For automated cross-verification, the ID numbers of individuals must be collected and available in all information systems, or a business intelligence algorithm to match individual characteristics, such as name, age, gender, address, or GPS coordinates must be used.

- Ensuring that different administrative databases can communicate, which includes the harmonization of technology, language and coding, and a data field dictionary, is essential when the goal is to create interoperability[19] and data integration[20] to improve the quality of means testing. The level of interoperability and data integration can be different according to the organizational needs, but countries must also have proper data-sharing protocols and good data protection in place. This harmonization allows different systems to interact, to validate self-declared information and generate other measurements that program administrators need to improve accuracy in the selection of beneficiaries.
- Regulatory frameworks and protocols for data exchange, privacy, and confidentiality of information must be in place.
- Convenient on-demand application—virtually (online or by phone) or at local centers with dedicated and trained staff to collect information from the applicant or to trigger the interoperable system of data exchange in centers must be established at least at the municipality level.

Due to their data intensity and requirements of interoperability, means tests are common in high-income countries where the economies are largely formal and extensive government databases allow for verification of incomes. The following are some examples of how they work:

- The United States defines eligibility for most social programs through an assessment of household income and selected assets, with rigorous verification to improve targeting accuracy. The United States uses social security numbers and state-level IDs to integrate information systems. Most of the registration, database management, and eligibility decisions are decentralized to the state and/or municipal (county) level, with federal oversight and fraud control for federally funded programs. For federal programs, the main program rules are set by the federal government, but there can be some room for flexibility at the state level. For example, rules set by the federal government restrict eligibility for SNAP to those with gross incomes up to 130 percent of the federal poverty line. However, in Alaska and Hawaii, benefit levels and income eligibility requirements are higher, and states can have different eligibility levels due to the exception of broad-based categorical eligibility.[21]
- The Portuguese cash transfer program for the extreme poor, Rendimento Social de Inserção, is offered to families with total monthly income below a threshold that varies according to the household size and demographic characteristics.[22] The total income is defined as the sum of verifiable income from salaries, pensions, housing subsidies, and other social programs, as well as an estimate of property income and capital income. For property income, two estimates are added to the total income: (1) 5 percent of the difference between the property value and €193,005

(US$216,860) if the property value is greater than €193,005, and (2) 5 percent of any other property. For capital income, the estimation adds a twelfth of the maximum amount between the value of income from earned capital (interest on bank deposits, stock dividends, or income from other financial assets) and 5 percent of the total value of capital income (as of December 31 of the previous year; bank account balances, shares, bonds, or other financial assets). Moreover, the program is not offered to applicants with a total value of capital income from bank account balances, shares, investments, and other financial assets that is greater than €26,145 (US$29,050).

- The Greek Social Solidarity Income program uses an income assessment based on declared income during the application stage but verified through interoperability of the systems used by different government agencies, which is made possible by the uniqueness of the applicant's social security number. Once completed, the registration form automatically pre-fills the fields in the form using the information stored with the tax authority. This information includes the name of the applicant, date of birth, family composition, and previous year's household income based on the applicant's tax declaration. Moreover, the Unified System of Social Insurance allows cross-verification of the employment status of all working-age adults in the household and any pension and/or benefits received, and an asset filter is applied.[23] Eligibility for the Social Solidarity Income program also includes an asset test.

Even in countries with significant informality and limited possibilities for verification of declared incomes, means tests are sometimes used, sometimes in combination with other methods.[24] In China, the *Dibao* program operates as a means-tested guaranteed minimum income. Early in the program's history, verification was not done with the sort of national income, social security, or property records that are commonly used in Europe, but rather through recourse to the information available to local government cadres and community members. More recently, information systems and accountability measures are being developed with more central frameworks (Golan, Sicular, and Umapathi 2014; World Bank 2021a).

In Chile, the launch of the Social Household Registry has marked a shift in the targeting method across all social protection programs, from PMT and HMT to a means test. The reform has not required collecting new information from applicants; it involves only accessing and using existing information from administrative databases (tax records, wages, social security contributions, health insurance [public and private] contributions, unemployment insurance, pensions [contributory and noncontributory], education records,[25] real estate, and vehicles). The Social Household Registry uses administrative data on formal sector incomes, complemented by self-reported informal incomes, to construct an indicator of

household income. The composition of the household is self-declared but interoperable with the national ID to guard against false or duplicate IDs.[26] The household income is then transformed into per capita equivalent form using a set of normative needs indices for different household members. Then, households are classified into seven socioeconomic groups, corresponding to the poorest 40 percent of the population (group 1) and six other groups corresponding to deciles five to ten. The first group was left deliberately as the largest, given that formal income information alone was not sufficient to rank households into deciles one to four. If a large share of household income is derived from informal sources and self-reported, the Social Household Registry validates the income information with an assessment of household means. The means test measures the possessions of five categories of means or expenses: cars, real estate, school tuition and fees, cost of health insurance, and the balances in pension accounts. If the household is ranked high on two or more means, its socioeconomic group is increased.

The Social Household Registry's new means test is used to prioritize all the social assistance benefits and services in Chile. A total of 80 programs use the means test/socioeconomic groups for eligibility. Some of these programs use only socioeconomic groups. This is the case, for example, of the noncontributory child grant, which is a child allowance for children of informal sector families in the poorest six deciles. Should a program target a group that is smaller than the 40 percent poorest, additional selection criteria are considered. For example, the Securities and Opportunities program is targeted toward extremely poor families and serves only a predefined quota of beneficiaries in each district or commune. For the selection of the potential beneficiaries, the households in the Social Household Registry are ranked by their specific income per adult equivalent, which is then compared with an income threshold equal to the extreme poverty line. If the quota is smaller than the resulting caseload, the program further prioritizes families with children.

Along the same lines but with more detailed verification, Brazil combines its means-test approach with geographic targeting and uses the social registry and related mechanisms of the delivery system for beneficiary selection for a large range for social programs. The Brazilian system was initially based on self-reported income with little verification when it was established in 2001 (the Bolsa Escola, Bolsa Alimentação, Cartão Alimentação, and Vale Gás programs were later merged under the Bolsa Familia program). Geographical targeting was used to distribute quotas of caseloads of participants to municipalities for budget rationed programs. Since 2005, the country has invested heavily in the Unified Social Assistance System, which created the Social Assistance Reference Centers at the local level. The main functions of the Social Assistance Reference

Centers are to assist people benefiting from Unified Social Assistance System services and to help them access other services provided under other sectoral policies. A Social Assistance Reference Center is equipped with the human resources and infrastructure needed to establish local partnerships, as well as to mobilize and inform communities around the importance of investing in human capital to increase their productivity and how to access services, programs, projects, and social assistance benefits. At a Social Assistance Reference Center, trained staff enter people's initial contact details into the Social Assistance System. During the face-to-face sessions, which collect the self-declaration of socioeconomic conditions in the Cadastro Único, people are informed about verification measures such that false declarations will lead households to a suspension of benefits as well as penalties that would block any member of the family from receiving assistance. Moreover, Brazil has always used cadastral surveys, spot checks, and audits financed at the federal level to assess the quality of self-declared data in the Cadastro Único; people's trust and confidence in the staff at a Social Assistance Reference Center is also believed to reduce bias and underdeclaration of income.

Since 2014/15, Brazil has performed partial verifications of income declarations in the Cadastro Único through data sharing with several government agencies. This requires different identification numbers for each household member, including the tax ID, national ID, social security number, and labor card number,[27] as Brazil still does not have an official unique digital national ID. After data entry, a cross-verification process involving different information systems runs simultaneously to compare declared income with tax records or the Relação Anual de Informações Sociais (RAIS) or formal employers information system (Cadastro Geral de Empregados e Desempregados [CAGED]) and the National Social Security system. This interoperability is recent and still partial due to the high informality levels in Brazil. However, several other processes, such as spot checks and random home visits, are also part of regular monitoring. The 2019 audit of social programs done by the Federal Court of Accounts (Tribunal de Contas da União) found R$2.2 billion (US$556 million)[28] in suspicious transactions of R$55.6 billion (US$14.1 billion) benefits paid, or less than 5 percent.[29] A total of 449,000 benefits were considered suspicious and 65 percent of those were receiving social assistance, which required the administrators to investigate the cases. If all those on social assistance were related to Bolsa Familia, the inclusion errors would be equivalent to 2 percent, as the program reaches more than 13.5 million households.

Means testing benefits from technology and government measures that are increasing digitization and e-governance in many countries. As technology improves and systems integration and interoperability allow the provision of better service delivery, better data quality, and better

information in general, more countries have/will have the minimum conditions to start planning the transition to use at least partially verified means testing as a targeting method. As chapter 5 discusses, some countries are moving in this direction. Investments in improving systems, improving their interoperability, and building the capacities for client interface in-person and remotely can speed the transition. Thus, means testing is both the gold standard and a method that is increasingly feasible (to one degree or another) in many countries.

Key Elements for Hybrid Means Testing

When much but not all of a family's or individual's income and assets can be observed and verified against independent sources, programs can impute or predict the remainder using an HMT. This solves a common dilemma for the administrators of social programs: placing on households unrealistic expectations of being able to know/disclose their welfare or dealing with large measurement error in assessing household income. Officials may ask the households to report all their incomes, knowing that informal incomes may be underreported and thus pushing some beneficiaries into committing fraud; or they can impute this income based on some verifiable information, such as asset ownership or the branch of activity of an informal worker. When imputing some values, the means testing becomes an HMT. Relatively few developing countries have adequate formality and information sources to conduct fully verified means tests that are valid over the entire population. However, many countries have only moderate informality and well-developed databases covering the formal part of the economy. HMT is designed for such circumstances, as it takes advantage of the information available on formal incomes in administrative databases and imputes some of what is not.

To assess whether to use an HMT, Tesliuc et al. (2014) suggest classifying income based on the country's administrative capacity level as follows:

- Easy-to-verify incomes are those from formal employment, earnings associated with formal entrepreneurship or asset ownership, and social protection transfers. They typically include the following:
  - Wages earned in the formal sector (subject to social security)
  - Nonwage benefits earned in the formal sector (bonuses and so forth)
  - Social transfers or social assistance (unemployment allowance, veterans' allowance, and so forth)
  - Retirement pensions
  - Dividends, interest received, and so forth.

- Hard-to-verify incomes are those from the informal sector: cash and in-kind income from agriculture, income from employment in the informal sector, capital gains, remittances, income from informal agreements for renting or leasing land or houses, and so forth. They typically include the following:
  - Gains from an informal and/or occasional secondary activity
  - Wages earned in the informal sector (not covered by social security)
  - Gains from agricultural activity.

Whether a country can achieve high accuracy using a means test (with informal income excluded) or an HMT depends on the share of informal income in total income and its variability. This is illustrated by a tale of two countries, Bulgaria and the Kyrgyz Republic, in figure 6.1. When the share of hard-to-verify income is relatively small and grows monotonically with

Figure 6.1   Distributions of Easy- and Hard-to-Verify Income in Bulgaria and the Kyrgyz Republic

a. Bulgaria: Share of hard-to-verify income, rural areas

b. Kyrgyz Republic: Share of hard-to-verify income, rural areas

c. Bulgaria: Composition of full income, rural areas

d. Kyrgyz Republic: Composition of full income, rural areas

● Hard-to-verify incomes   ● Easy-to-verify incomes   ── Eligibility threshold
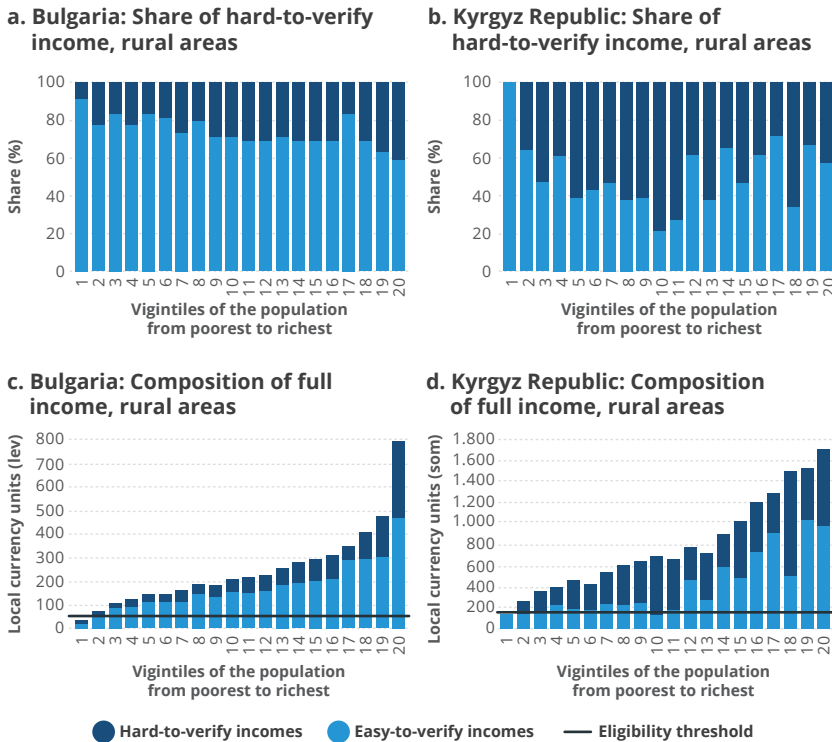
Source: Tesliuc et al. 2014.

income, the ranking of households based on verifiable income provides a good approximation of the distribution of the full income. In this case, means testing is recommended as the risk of inclusion errors is relatively low, as in the case of Bulgaria. However, when the share of hard-to-verify income is high and not positively correlated with income, the risk of inclusion errors (or rank reversal) using means testing is higher, as in the case of the Kyrgyz Republic, and HMT can then be recommended.

Figure 6.2 illustrates graphically how accounting for hard-to-verify income would reduce the inclusion error in an income-targeted program, compared with the situation where such income is not considered or is mandated by the program but not reported. If all income could be observed, the number of eligible applicants would be the area 0A in the figure and inclusion and exclusion errors would be zero. Not considering the hard-to-verify income increases the number of accepted applications from 0A to 0B. The segment AB represents inclusion error. Including the estimated hard-to-verify (presumptive) income reduces the number of accepted applications to 0C, thus eliminating some of the inclusion error (the population segment CB).

Figure 6.2   Income Test under the Hybrid Means Testing Approach
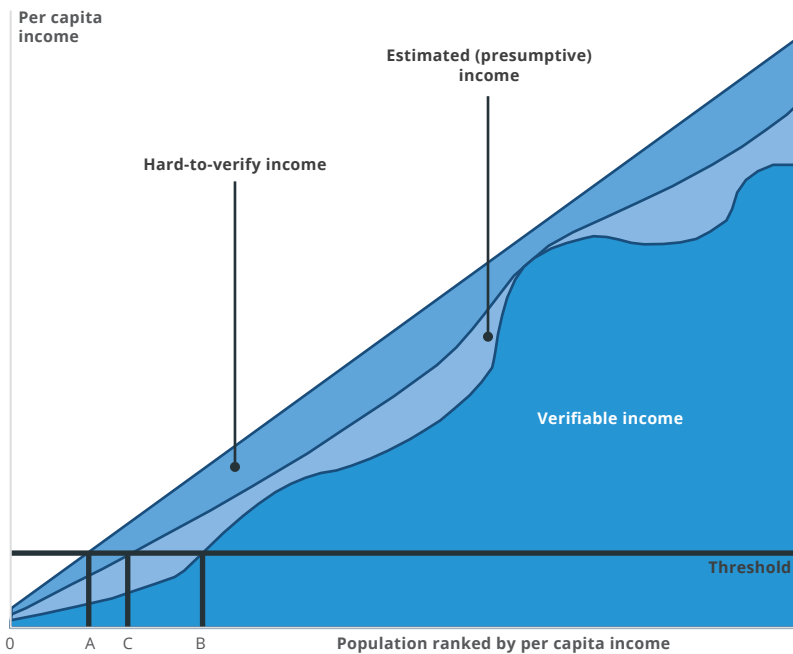
Source: Tesliuc, Leite, and Petrina 2009.

There is no single method for estimating the hard-to verify income. The context of each country largely determines the choice of the method. The administration could use labor market surveys, individual interviews, or the subjective evaluation of experts to determine the level of income to be attributed to each informal activity. The imputation method or values imputed can also vary from one region to another to account for local variation.

Imputation of expected income from hard-to-verify sources is at the heart of HMT. Those designing HMT tend to focus on some of the largest informal employment pools or branches of activity and develop simple rules for estimating that income. In the Europe and Central Asia region, most informal employment occurs in the agriculture and construction sectors. In this case, social programs can develop simple imputation rules for income in these sectors, rather than using regression models. An analyst may use some of the simple methods to impute informal income to add to formal income measures; more complex approaches may be useful where informal employment is spread across many sectors and thus more difficult to estimate.

There are significant inherent difficulties in measuring the agricultural income of small farmers whose households are dual production and consumption units.

- Small farmers often operate as dual production and consumption units, without keeping separate accounts for what is used in the production process and what is consumed. A proper accounting of the value added generated by the farming household will separate the production account and estimate the value added produced as the sum of the (often implicit) labor earnings, imputed rent from the land owned, and residual profits. However, this is difficult and rarely done in practice. Not all the revenues or expenditures on inputs (for example, labor, equipment, and fertilizers) are monetized because a part of the production is consumed by the household, and some inputs are produced or supplied by the household (for example, fodder and some labor). Small farmers, often at risk of poverty, consume a larger share of their production themselves, bypassing markets and making the valuation of their outputs and inputs a complex task. This portion of farm income appears in the "consumption out of own production" estimates in household surveys.
- Farm income can be measured only at the end of an agricultural season, which is often over a calendar year and thus much longer than in other economic branches where such estimates could be generated monthly. In the Europe and Central Asia region, the agricultural season for crop production is typically a year and for livestock production several months, depending on the type of livestock. Over the agricultural season, expenses are incurred during the planting or breeding time, whereas

revenues are collected only after harvest. Only then can farmers determine the net revenues after costs and calculate their profit. In contrast, day laborers or formal employees would know their income (wage) at the end of the day or month, respectively.

Owner-operated small enterprises in several nonagricultural branches of activity also have joint production and consumption accounts, which again makes income measurement hard. A taxi driver may similarly mix accounts—using the vehicle for family as well as client trips, taking repair money from savings, and taking gas or lunch money from daily earnings before bringing them home as earnings.

Acknowledging that it is difficult to measure incomes from small farming, many social programs in Eastern and Central Europe or the former Soviet Union have developed simple, practical approaches to estimation based on asset ownership and their estimated returns. For verification purposes, and to mitigate the risk that applicants do not report asset ownership, the programs rely on land and/or livestock registries. The registries are also used to validate changes in asset ownership, use, quantity, or quality. The quality of the income imputation depends on the quality of the information in the registry. Among the factors that would improve the precision of the imputation are the availability of information on the quantity and quality of the asset, timely information on the owner and user, and the frequency with which the information is updated.

A simple imputation approach has been and, in some cases, still is used in Albania, Armenia, the Kyrgyz Republic, Lithuania, and Romania. In the case of farm income, imputed income is estimated based on the type, location, and quality of the land. Income from livestock production is estimated based on simple farm models, with the income coefficients often estimated in consultation with agricultural agencies, research institutes, or ministries. These institutions also have a role in validating the estimates. Until 2018, Albania's Ndhima Ekonomike[30] offered a textbook example of the imputation of agricultural income. Imputation for land was done based on revenue coefficients that vary according to the type of land and geographic characteristics (table 6.1). Presumed income from livestock production was differentiated by the type of livestock and three geographical zones. A similar approach applies in the Kyrgyz Republic, where individual revenue coefficients vary by location (more than 480 locations) and type of land (personal plot versus agricultural land, irrigated versus nonirrigated) (Government of Kyrgyz Republic 2018; OECD 2018). The possession of livestock works as an exclusion filter, whereby each animal is converted to a number of notional units and all such units are then summed up and divided by the number of household members. Program eligibility requires that the number of such units per capita does not exceed a predefined program threshold.

Table 6.1   Albania: Imputation of Farm Income

Presumptive income rules based on land ownership—land categories and revenue coefficients, Albania									
	Land categories								
Item	I	II	III	IV	V	VI	VII	VIII	IX–X
Revenue coefficient lek/m² per year	9	8	7	7	5	5	4	4	3

Livestock unit conversions and revenue categories	
Zone	**Coefficient**
Lowlands	Revenues from 1 cow = 15 sheep or goats = 3 swine = 5 piglets = 20 beehives = lek 22,500/year
Hills	Revenues from 1 cow = 12 sheep or goats = 3 swine = 5 piglets = 20 beehives = lek 18,000/year
Mountains	Revenues from 1 cow = 10 sheep or goats = 3 swine = 5 piglets = 20 beehives = lek 13,000/year

Source: Council of Ministers, Albania 2005.
Note: US$1 = lek$102.93 in 2005.

A more complex imputation method, yet still simple enough, relies on more detailed and precise information on returns to land and livestock, using agricultural registries. In Romania, from 2014 onward, researchers refined their estimation of income from agriculture, which was previously estimated using simple productivity coefficients as in Albania, based on the survey of small farm holdings harmonized across all EU countries. The survey allowed yearly estimation of gross annual margins for different types of harvests or livestock. These coefficients (table 6.2), which are representative of the earning potential of the small farms (typically informal sector farms), were then applied to the land owned (or leased) by each farmer and its stock of livestock (available in the agricultural land registry) to generate the estimated agricultural income for each household in the program. The program procedures allow the adjustment of these numbers in case of unexpected events. If the farmer experiences a shock and the agricultural production is lost in whole or in part, an agricultural extension worker certifies the loss of income and the presumptive income is excluded from the calculation of the total income of the assistance unit. This method allows a more precise estimate of farm income, based on the ownership of key agricultural assets such as land and livestock, at more frequent intervals.

Another example of simple income imputation is for unskilled seasonal work or day jobs, or for informal workers in key occupations/sectors. For example, in addition to the hard-to-verify income from land and livestock, for beneficiaries capable of working, Albania, Romania, and Uzbekistan also impute presumed earnings from seasonal occasional work—at local market wage rates—for a given number of days per month during the

Table 6.2   Romania: Asset-to-Income Conversion Coefficients

Gross annual margin

	Type of crop/harvest	(Euro/year/ha)
1	Common wheat	170
2	Durum wheat	74
3	Rye	75
4	Barley	112
5	Oat	90
6	Corn	237
7	Rice	52
8	Other grain	198
9	Potatoes	1051
10	Sugar beet, without seeds	720
11	Rape	195
12	Sunflower	173
13	Soy	244
	...	
42	Poor pastures infertile or uncultivated	14
43	Orchards of apple, pear	836
44	Orchards of plum, peach, apricot, cherry	836
45	Walnut orchards, hazelnut, chestnut	269
46	Orchards of currant, fig, raspberry	269
47	Grapes for quality wine	672
48	Grapes for table wine	604

Gross annual margin

	Type of livestock	(Euro/year/ha)
1	Horses	57
2	Calves	138
3	Cattle 1 to 2 years	59
4	Breeding heifers	87
5	Cattle for fattening	98
6	Dairy cows	391
7	Rabbits, for females, hoeing	25
8	Goats, for females, hoeing	57
9	Sheep	3
10	Piglets	24
11	Sows for breeding	17
12	Pigs for fattening	115
13	Broilers	9
14	Hens	5
15	Other birds	11

Source: Tesliuc et al (2014).
Note: ha = hectare.

agricultural season. Similarly, allocating a lump sum income equivalent to the average earnings observed (or known) for the main employment occupation and sectors by regions allows a simple estimation of informal income. Average income can be estimated by age, gender, occupation, sector, and region from household or labor force surveys and applied to the applicants.

In the cases where informal income is derived from many types of sectors and occupations that are difficult to account, estimate, and verify, a more complex variant of the HMT method is occasionally applied. At the most general level, under this variant, applicants are subject to two welfare tests, an HMT and a PMT. Those whose HMT income is below the program threshold and whose PMT score is below the corresponding threshold are eligible for the program. In the first step, using a means test/HMT eliminates inclusion error of people with observable or easily imputed incomes. The reason for the second step is concern that the estimation of informal income in the HMT may be downward biased compared with the true informal income. This could occur because some informal activities are not covered or estimated under the HMT, or because the level of imputed income is set at a lower level than the true value (to encourage self-reporting, as is typically done with assets or occupations taxed based on presumptive income). To reduce inclusion error from this source, the remaining applicants are subject to a full PMT. From the implementation perspective, this variant supposes the development of both HMT and PMT testing capacity. In design and implementation, this variant is more complex than the preceding one for the administration and the beneficiaries, although data sharing and interoperability are reducing these complexities. It is also less transparent than simple income imputation, for reasons related to PMT targeting, which are discussed in the next section.

Two-step HMT eligibility was pioneered in Moldova (Carraro 2014). The Moldovan guaranteed minimum income program (Ajutor Social) uses HMT to provide cash income to eligible families. As with any guaranteed minimum income program, the benefit amount is the difference between the household-specific minimum income threshold and the actual household income. To qualify for the Ajutor Social benefit, the applicants must pass an HMT: an income test based on verifying formal income using governmental databases, plus income imputations for other income sources. Types of income that are not easily verifiable—such as agricultural income—are imputed. Agricultural income is estimated based on the amount of land owned by the household, considering the type of plot (whether the land is close to home or farmland), the different agricultural zones, the fertility of the soil, as well as whether the household has some livestock. The value of the land is assessed at the level of the locality in the land register along with the lot size. The estimated amount of net income that people derive from a hectare of land is updated once a year based on estimates provided

by the Ministry of Agriculture and the National Bureau of Statistics. For applicants who pass the first filter, a PMT is used. For each applicant family, a score is determined based on age, education, disability condition of the family members, and ownership of assets, durable goods, or consumption items, such as gas consumption. To be eligible for the program, the household has to meet both the income test (formal income plus estimated informal income lower than the program eligibility threshold) and the PMT (PMT score below a predefined threshold).

Turkey is unusual in bringing to bear consumption information in a two-step HMT[31] (Ortakaya 2020, 2021; Turkey MFSP and World Bank 2016). The estimated income in Turkey's HMT includes two components: formal income from wages, social protection transfers, alimony, interest on deposits/savings, or rental income, which is recovered as such from other administrative databases; and estimated income, called equivalent rental rates and defined by Turkey's Ministry of Finance, for ownership and use of additional dwellings owned, business premises (as example, shops), urban or agricultural land, passenger car or cars, commercial or agricultural vehicles, and livestock. The rental income of the owner-occupied house is disregarded. Households are also asked to self-report their monthly consumption per capita for several expense categories (food, clothing, rent, health, education, transportation, entertainment, and tobacco consumption). If the estimated income falls short of the expenses, the level of expenses is considered. The HMT is used to differentiate the waivers or subsidies to health insurance premia for adults in the informal sector and their dependents, as part of universal health coverage. If the estimated income is less than one-third of the gross minimum wage, the premia is waived; if the estimated income falls between one-third and two times the gross minimum wage, the premia is subsidized; and if the estimated earnings are greater than two times the gross minimum wage, the applicant must pay the full premia, similar to employees in the formal sector.

In practice, the combination of easy-to-verify income and estimated hard-to-verify income generates three stylized situations with different implications for errors: households that earn (1) only easy-to-verify income, (2) only estimated hard-to-verify income, or (3) both.

- For the segment of households whose income is derived in full or in large part from formal income sources, the HMT method is as accurate as means testing. For this group of applicants, the HMT method matches their actual welfare level.
- For the segment of the population that derives income mostly from informal or hard-to-verify sources, the imputation method results in inclusion and exclusion errors. The level of error is proportional to the dispersion of such earnings around their mean.

- For the segment of the population that derives income from both formal and informal sources, the method could approximate quite well the true welfare level on average, because formal incomes are on average higher than informal ones.

The overall performance of the method depends on the relative shares of the three types of households. Accuracy will be higher if most of the households are in the first and third categories. The method will not work well if households are broadly separated into the first and second categories and the share of the second category is large. In this case, HMT would not be appropriate, although PMT may be more so. To select between HMT and PMT, a simulation of the two targeting methods based on household survey data is recommended. The example of Algeria in box 6.3 illustrates the value added of this approach.

BOX 6.3

Simulating the Accuracy of the Hybrid Means Testing and Proxy Means Testing Models in Algeria

In Algeria, formal employment represents half of total employment and formal incomes are about two-thirds of total household income. Algeria carried out a simulation to compare the performance of a hybrid means testing (HMT) model and a proxy means testing (PMT) model in 2019. For the HMT, the simulation assumed observability of formal wages and social protection transfers. The main sources of informal income were imputed based on two simple scenarios. Under the first scenario, the median earnings by branch of activity and region were imputed to all potential applicants. Under the second scenario, 75 percent of the median value was taken instead to reduce exclusion error and encourage reporting. The program definition of income was the sum of the two components. The PMT model estimated household consumption using typical covariates (household size and composition, education, employment and sector of activity of the adults, endowment of durables, dwelling characteristics, and region).

The assessment suggested that the HMT targeting model would more accurately identify low-income households (for several thresholds, such as the 10, 20, or 30 percent poorest) than a PMT model. The simulations led the government of Algeria to decide to develop a national database of formal incomes and assets, as well as implement other incremental reforms to improve the quality, timeliness, and availability of administrative data.

The HMT would work well in countries where (1) most informal income is concentrated in a few economic sectors; (2) the variation in earning rates per worker is relatively low; and (3) there is a relatively simple way to estimate the average level of informal earnings per sector, as the average returns of a productive asset (for example, land, livestock, and equipment) per month, or as an average per occupation type (for example, average taxi earnings in a given town). Among the key advantages of this approach are transparency and simplicity. Beneficiaries can easily understand the program eligibility criteria and determine on their own if they are eligible or not for a program and what level of benefit they might receive. This increases the political acceptability of the program among beneficiaries, social assistance staff, and the population at large.

In considering whether to use HMT, the program should also test the quality of the administrative data systems that would support it. This assessment will tell program administrators whether what was simulated based on household survey data is supported by the data infrastructure of the country, and whether HMT could work in practice.

The ideal situation for HMT to work well is one in which formal earnings and social protection transfers (including pensions) are reported regularly, in full (in terms of coverage of the income recipient and income level), with this information kept in databases that are interoperable and can exchange the relevant information on an as-needed basis. Apart from the technical prerequisites, access to information is also regulated by the data privacy regulations in force. For example, in many countries, information on savings and interest income cannot be accessed due to privacy regulations. Some countries have laws that restrict the use of tax record data. Moreover, the frequency of income reporting also matters for the ability of HMT to respond rapidly to changing circumstances. Countries where the social security agency collects earnings information annually could implement social protection programs using HMT that are less adaptive than those in countries where earnings are reported monthly.

A few factors can reduce the accuracy of a would-be HMT system. In some countries, certain categories of formal employees do not have individual earnings records in the social security system (for example, for all public sector employees or those working in national security, the army, police, or magistrates/judiciary). Social security records sometimes cover only the base wage and do not record elements such as premia for hardship, merit, and so forth. To encourage contributions to social security among higher-income earners, some legislation allows capping the earnings they report, with a similar cap on the maximum pension and related benefits. This will diminish the accuracy of the income information available in the social security database. At the other end of the income reporting spectrum, some social security administrations use minimum imputed

income wherever the contributor's self-declared income is below a certain threshold. Consequently, in all these cases, the accuracy of HMT will likely diminish.[32] The example of Tunisia illustrates the value added of carrying out an administrative data assessment as part of the program design (box 6.4).

BOX 6.4

Administrative Data Preconditions for Developing a Hybrid Means Testing Targeting System: The Case of Tunisia

Tunisia seemed to be ready for the implementation of a hybrid means testing (HMT) system. Formal employment represented about two-thirds of total employment, the share of formal earnings in total earnings was about 80 percent, and half of the elderly were covered by the pension system. In 2018, the country embarked on a series of technical studies to assess the feasibility of introducing an HMT targeting system. On the design side, the exercise has been hampered by the lack of a household survey that collects information on income and, more generally, the lack of a survey that collects comprehensive information on income and assets. To assess the availability and quality of the administrative data, a technical assessment was conducted (CRES and World Bank 2018), the results of which are summarized in table B6.4.1. Information on the quality of the Social Registry data was recently updated during the preparation of the Tunisia COVID-19 Social Protection Emergency response support project.

The assessment confirmed the feasibility of introducing an HMT targeting system in Tunisia because of the quality of data, formal income coverage, and interoperability within and beyond the social protection systems. The National Pension and Social Security Fund, the agency that covers public sector employees (a fifth of total employment), keeps individual income records for the civil service (80 percent), but only the overall wage bill for publicly owned enterprises (20 percent). However, the individual records of the employees in publicly owned enterprises are at the enterprise level and could be mobilized in the future. In the case of the incomes declared to the National Social Security Fund, the pension agency for the private sector, income records were assessed as accurate in terms of coverage of affiliated

continued next page

BOX 6.4 *(continued)*

Table B6.4.1   Summary of the Availability and Quality of Administrative Data in Tunisia, 2019

Administrative data and administrator	Overall	Coverage	Quality	In electronic form?	Inter-operability
Wage earnings and assimilated					
Taxpayers (Direction Générale des Impôts)	A reliable source, but not accessible	(half)	(quality high)	(electronic high)	(star high)
Social security contributions and reported earnings (CNSS, CNRPS)	Good coverage and quality for wage earners, partial info. on self-entrepreneurs	(half) (part) 3 million contributors	(quality medium)	(electronic high)	(star medium)
Social protection transfers					
Pensions (CNSS, CNRPS)	Reliable data for 50% of the population 60 years old or more	(full) 1 million pensioners	(quality medium)	(electronic high)	(star medium)
Social assistance transfers (MAS)	AMEN Social program: (1) quality info. for the permanent cash transfer (PNAFN), (2) less so for subsidized health card (AMG2)	(full) (half)	(quality medium) (quality low)	(diamond) (triangle)	(star medium) (star medium)
Agriculture income					
Ministry of Agriculture	Information at the level of the Ministry of Agriculture, not centralized, partial and in paper form	(low)	(quality low)	(diamond)	(star low)
Assets and capital					
Land ownership title (Conservation de la propriété foncière)	Kept in paper form in regional offices	(half)	(quality low)	(triangle)	(star medium)
Car title/carte grise (Agence Tunisienne des Transports Terrestre)	Information on 2 million vehicles and their characteristics	(full) 3 million vehicles	(quality low)	(triangle)	(star high)

Legend

	Coverage	Data quality	In electronic form?	Interoperability
Low	(light circle)	(quality low)	(diamond)	(star outline)
Medium	(half circle)	(quality medium)	(triangle)	(star half)
High	(full circle)	(quality high)	(full circle)	(star full)

Source: Centre des Recherches et des études Sociales (CRES) and World Bank 2018.

Note: AMEN = safety (in Arabic); AMG2=Assistance médicale gratuite 2; CNRPS = caisse nationale de retraite et de prévoyance sociale/social fund for the public sector; CNSS = caisse nationale de sécurité sociale/social fund for the private sector; PNAFN = programme national d'aide aux familles nécessiteuses; MAS = ministère des affaires sociales/ministry of social affairs.

continued next page

BOX 6.4 *(continued)*

workers, but they were less reliable for the level of income. By triangulating data from survey and administrative sources, CRES and World Bank (2018) estimate that there is potential underreporting of one-quarter of the wage bill, especially among nonwage contributors. Informal agricultural and nonagricultural incomes, as well as data on interest, capital income, and tax records, were confirmed as hard to verify. Finally, the report recommended steps to improve the accessibility and quality of data (removal of legal barriers in accessing administrative data while respecting the privacy of the provider, improvement in wage reporting, and, in some cases, moving from paper-based to electronic records). With these improvements, HMT could be implemented in the future for selecting beneficiaries for social programs in Tunisia. The country's ongoing efforts include interoperability and a new household survey, which collected income data for the first time in Tunisia.

Source: CRES and World Bank (2018).

Key Elements for PMT Methods: Traditional Models, Processes, and Machine Learning

PMT is an inference-based assessment of household income or consumption, which is used when means testing or HMT is not available. The technical details require greater elaboration compared with other methods. PMT uses statistical methods to estimate a household's level of income or consumption or its eligibility on a monetary welfare metric. Given that PMT estimates welfare indirectly based on other observable household characteristics, it is more complex and opaque than other means-testing assessments. At the same time, with many developing countries having large informal populations and no direct data on their income or consumption, PMT is a widespread targeting mechanism. Much of the remainder of this chapter is devoted to the technical details on how best to implement this imperfect but commonly used alternative.

PMT was initially developed in Latin American countries, beginning with Chile's Ficha CAS (Comité de Asistencia Social [Social Assistance Committee]) in the early 1980s. Other early and iconic examples are Colombia's System for the Selection of Beneficiaries for Social Programs (SISBEN), which was launched in 1994; Costa Rica's Sistema de Información de la Población Objetivo (SIPO), which was inaugurated in 1999; and the registry for

Mexico's PROGRESA-Oportunidades-Prospera program, which operated from 1997 to 2019. PMT was developed to identify poor households in the context of high levels of informality, inequality, and poverty in the region. The method spread to many other countries that desired to focus their programs on the poor or vulnerable, mostly countries with significant rates of informality but often much lower levels of inequality.

A PMT is a model for translating readily observable household, community, and regional characteristics (explanatory variables) into an estimate of household consumption or income based on any of several forms of statistical modeling. The predicted weights obtained from the model are then applied to the information for program recipients to estimate household welfare and thus determine program eligibility. Household-specific information is usually obtained from households self-reporting in an office interview or home visit. Often the information is on characteristics that are easily observable, to prevent error and fraud. Location-specific characteristics obtained from administrative data or other systems—such as early warning systems, census data, poverty/vulnerability maps, or even for other sensing data (see the subsection on data-related limitations and considerations)—may be included to improve the estimations. Consumption or income can be estimated in PMT, but it is advisable to use the metric that is used by the country to measure and determine official poverty. The modeling is predictive only, with no pretense of providing causation (box 6.5).[33]

A handful of programs use scoring formulae that are not derived based on statistical modeling but on the expert opinion of program designers or the experience of social workers and their perception of the factors that are associated with poverty. Some programs start like this and then migrate to a statistically derived formula once the relevant microdata are collected and analyzed. The Kenyan Cash Transfer to Orphans and Vulnerable Children program started by using a poverty test that was based on 17 binary questions (yes/no questions recorded during application) in addition to three other methods: categorical targeting (to determine whether a child was an orphan, CBT for preselection of potential beneficiaries, and geographical targeting due to budgetary constraints. Any household exhibiting eight or more "yes" answers for these questions was classified as poor. After the evaluations[34] showed limited efficiency in finding the poor, the program migrated to a full-fledged PMT scoring formula to replace the poverty test. The Moroccan Medical Assistance Plan is a health insurance waiver program that serves the poorest quintile of the population. It operated for a decade based on two ad hoc scoring formulae, for rural and urban areas, respectively, and will be replaced by a full-fledged PMT and social registry in the near future.[35] In all these cases, the factor that triggered the change from an ad hoc formula to an analytically derived scoring model was a comparison between

BOX 6.5

PMTs Are a Predictive Model Exercise, Not a Causal Effect One

In general, statistical models are powerful tools for causal explanations, prediction, and description of data. A great deal of research/econometric analysis uses statistical modeling to test causal claims. In a regression, the causal claim follows a simple structure in which each of the covariates (called independent variables) is assumed to have a causal influence (regression coefficient) on the dependent variable (for example, income or consumption per capita). The models are based on the assumptions that covariates cannot have any causal influence on one another and there is no reciprocal causal influence from the dependent variable to any of the covariates.

In the targeting world, most of the time, similar models are used for their ability to predict what income or consumption would be when it is not measured. The inference is not causal but rather about association. Hence, strong underlying assumptions that are needed to determine causality are not needed or are incorporated in a less formal way. Consequently, the best model is not the one with high explanatory power or R^2, but the one with high predictive power, which is quite different.

Shmueli (2010) highlights the main differences between explanatory and predictive modeling. First, predictive modeling tends to have higher predictive accuracy than explanatory statistical models. Second, predictive models aim at (1) looking for association between the x (covariates) and y (dependent variable), (2) not having a requirement for direct interpretability in terms of the relationship between x and y, (3) having a forward-looking approach instead of testing an existent set of hypotheses, and (4) reducing at once the combination of bias (the result of misspecification of the model) and estimation variance (the result of using a sample). Addressing these points in predictive models translates into a different approach for selecting the covariates.

While building a model for proxy means testing (PMT), the aim is to find correlations and associations rather than to look for causal structure, endogeneity, or reverse causality. The main criteria for selecting the set of covariates are the quality of the association between them and the dependent variable, as well as preexisting knowledge of correlation/association that does not necessarily come from the data set but from other studies or local knowledge.[a] This procedure is different from explanatory models, where researchers must (1) only keep significant variables in the model, (2) address

BOX 6.5 *(continued)*

multicollinearity, (3) have clear/independent control variables, and (4) minimize endogeneity to address causality.

Finally, model selection in predictive modeling is not based on explanatory power—assessed using metrics computed as R^2-type values and the statistical significance of overall F-type statistics. The researcher can retain covariates that are statistically insignificant if the variable has importance for prediction.[b] The predictive power of models is measured by their capacity to predict an event using new data (Geisser 1975; Stone 1974) or carefully using the same data. Usually researchers focus on extracting holdout (a subsample from the same data) or pseudo-samples. In the targeting context, beyond measuring whether the average prediction and errors are acceptable overall, researchers must analyze the predictive power of the model for certain marginalized groups or groups that may be of interest for social policy. For example, good predictive power for the average income or poverty levels in region x does not guarantee that the same model would generate acceptable errors for households with elderly living alone, small households, or female heads of households.

The role of the researcher differs for traditional PMT and PMT that uses machine learning. Traditional PMT may be thought of as a study, not a technology, that requires good data mining and skilled researchers to make decisions, as it is important to understand historical data as well as existing external data to find patterns and behaviors. PMT using machine learning, in contrast, is a methodology for prediction that often uses artificial intelligence techniques. With these models, the algorithms are often given data and asked to process them without a predetermined set of rules and regulations. In this case, the systems adapt and learn as and when new data are added, without the need of being directly programmed, and without continuously addressing the discontinuity of loss functions.[c] Machine learning is data driven, and the problem to be solved needs to be precisely described to find the right algorithm, as once it is calibrated for a particular event, the model cannot be used for a different one. It must be reestimated, so the role of the researcher is important upfront in problem specification.

a. For example, in work on a PMT for a given country in 2008/09, the addition of the interaction between possession of camels and region x would improve the predictive power significantly, although the interaction was not significant according to the p-value. The rationale behind it was the local knowledge of the team that highlighted the main difference in welfare, which was not observed in the data due to the small sample size. Given the dryness of the area, subsistence farmers in region x owned camels for trading limited goods.

b. Shmueli (2010) explains that in medical research, a variable for smoking habits is often present in models for health conditions, whether it is statistically significant or not, and that sometimes exclusion of significant variables improves predictive performance.

c. See chapter 8 for a fuller explanation of using machine learning for targeting.

the incidence of beneficiaries in the current program and the simulated PMT model. The comparison showed that such a move would reduce the program's inclusion and exclusion errors. In other countries, the ad hoc formulae have passed the test of time. This is the case in Armenia, where eligibility for the two flagship antipoverty programs, the Family Benefit and Social Benefit programs, is based on a complex vulnerability score.[36] The incidence of program beneficiaries has been compared with an analytically derived PMT model, but the improvement in targeting accuracy was not considered to be enough to warrant a reform of the eligibility criteria and associated changes in the delivery system. The Kosovo Social Assistance Scheme shares a similar story.[37]

As PMT is an inference-based method, it contains statistical error, which makes it controversial. For example, Brown, Ravallion, and van de Walle (2016) simulate the performance of various PMT methods using data from nine African countries. They show how PMT helps filter out the nonpoor but excludes many poor people. Kidd (2011) and Kidd, Gelders, and Bailey-Athias (2017) stress that PMT has in-built design (statistical) errors and static instruments as well as some inevitable level of implementation errors. They argue that since errors can be high and the methods not well understood by applicants or communities, PMT can be perceived as arbitrary or a lottery. The works cited in this paragraph are among the more strident in tone in criticizing PMT, but everyone engaged in PMT work acknowledges the inherent imperfection in having to rely on statistical modeling rather than more precise measurement of welfare. But PMTs are not meant to be used where means testing or HMT is feasible. Many authors who reject PMT recommend geographic or demographic targeting as simple metrics of welfare, which can be thought of conceptually as single-variable PMTs and thus even less accurate, although they are simpler and more transparent. The prior chapters consider whether PMT is a suitable choice in a given setting and how to reduce implementation error, which are topics that are not rehashed in this chapter. The focus here is on how to reduce statistical error, although it cannot be eliminated.

This section reviews the main steps for developing PMT and the traditional data and models that underly it. PMT was first developed four decades ago and, while the basic policy problem to be solved is the same, the how-tos have evolved over time. The section is broken into several subsections: model choice, whether to use multiple models, how to choose explanatory variables, how to update the model and the social registry, and data-related limitations. All the discussion focuses on what have until relatively recently been the standard approaches to PMT, which are referred to as "traditional." A subsequent section considers how the advent of big data and machine learning might improve PMT.

Traditional PMT Models

PMT is fundamentally a question of inference. This subsection briefly surveys the traditional options for predicting income or consumption from available proxy data. PMT is the process of estimating unobservable household income or consumption from observable proxies that correlate well with it. This inference can be done on different bases. This subsection examines how proxies can be used with traditional approaches that rely on classical regression methods; more recent machine learning algorithms and big data–based proxies are examined later.

The PMT models with the lowest data requirements use principal component analysis. This method was made popular by Filmer and Pritchett (2001) and does not require a household survey with income or consumption as all other PMT methods do. It identifies linear combinations of variables measured in household surveys, which maximize the variation in the characteristics correlated with welfare across households. Usually, just the first principal component identified by this method is kept as a proxy for household welfare. Principal component analysis has significant limitations—the resulting score can be ranked but is difficult to interpret, standard inequality measures such as the Gini index cannot be calculated, and it cannot be compared with principal component analysis scores from other models—but it obviates the need for income or consumption surveys. This method can still be used—and is the basis for a household welfare proxy in the widespread Demographic and Health Surveys—but given the improvements in the availability of household surveys with income or consumption, a range of more accurate regression techniques are available for determining beneficiary eligibility.

Logit or probit models regress the binary status of a household (poor or nonpoor; eligible or not eligible) on the explanatory variables. Ordinary least squares (OLS) models regress household income or consumption on the explanatory variables. Both models are popular as they are more easily understood than principal component analysis and construct a direct probability of a household being poor (logit/probit) or a direct income or consumption measure (OLS). Both models are used in practice, but the main advantage of OLS is that the model coefficients are not generated relative to a fixed poverty line. With binary models, different versions must be estimated for different cutoff points (for example, at both the poverty and extreme poverty lines), meaning there is less flexibility to use the same scoring for different programs with different eligibility thresholds. Other models, such as generalized least squares and nonparametric models, are also used.[38]

The main weakness of OLS is that it assumes that the association between the explanatory and dependent variables is the same at all levels of the distribution. That is, it assumes that the regression coefficients are constant

across the population. Cook and Manning (2013)[39] highlight the need for thinking beyond the mean as the correlations and characteristics of the population between welfare and certain variables may be different for those at the bottom, middle, or top of the distribution. OLS models for targeting would then be inappropriate if the mean is not a good representation of the poor. In other words, if PMT attempts to identify those at the bottom of the distribution (the poor), why not estimate the model under the assumption of a population average effect for that income range while still estimating continuous welfare? Different approaches can be incorporated into OLS models for PMT, such as adding interactions to address this limitation.[40]

Koenker and Bassett's (1978) quantile regression method allows estimation of the coefficients in a direct and transparent way at the poverty rate/proposed eligibility threshold, which better matches the policy problem of targeting. A feature of quantile regressions is that they focus on limiting errors at the bottom end of the expenditure distribution by ensuring that the formula effectively models the expenditures of the poorest households, without caring about the imprecision in the formula above the poverty threshold needed for the program. This is an advantage if policy makers care about exclusion error more than inclusion error. A challenge is setting the right quantile. In practice, the first and second deciles or the first quartile is often used as this represents the program population of interest. Del Ninno and Mills (2015) suggest that when the poverty threshold lies far from the mean, or when between-cluster correlation is high, it may be more relevant to estimate PMT weights using quantile regressions. Brown, Ravallion, and van de Walle (2016) show that for nine African countries, quantile regression performs better than traditional models in most cases. In other words, quantile regression may be more appropriate for countries with high levels of poverty and low inequality, as the average population is not a good representation of the poorest; when much of the distribution is similarly poor, PMT models struggle to distinguish between more finely grained degrees of need.

Two alternative methods to quantile regression that aim to give more weight to the poor are poverty-weighted least squares and truncated regression. The first allows various weighting schemes (such as weighting equally for observations below the poverty line but giving zero weight to those above the line), while for the second, the sample is truncated for certain ranges (for example, the nonpoor). However, in practice, this does not improve the models as some variables are lost, which reduces accuracy. Quantile regression is considered more effective in certain contexts[41] because it weights different portions of the sample to generate the coefficient estimates, thus increasing the power to detect differences in the upper and lower tails.

With the advent of new computational algorithms, machine learning algorithms are being used to estimate PMT, using parametric and nonparametric models that are more computationally intense (for example, nonlinear models and tree-based models). A full treatment of machine learning algorithms is presented in the next section, but all the discussion below referring to regional models, choice of variables, updates, and so forth that are valid for traditional PMT and machine learning models.

Multiple Regional Models

An important policy consideration is whether to use a single national model or multiple regional models that are fine-tuned for different purposes. Different models for different regions are common. The values of different proxies can vary from place to place due to differences in such things as climate and preferences. For example, in rural areas, having livestock may indicate prosperity; but in urban areas, not having livestock does not necessarily mean that a household is poor. An air conditioner may be a marker of prosperity in a location with very hot summers but not somewhere with more mild temperatures. Thus, most countries use different models for different parts of the country. The extent to which this can be done depends on how much disaggregation the survey data allow. The most extreme example is Indonesia, where 500 different models were developed, one for each district. This is possible by pooling the large annual surveys from consecutive years. In simulations, the biggest gains came from pooling three years of data. In each additional disaggregation—from a single national model, to urban and rural, to provincial, to districts—accuracy improved, but the move to 500 district models from 71 provincial models[42] showed the greatest improvement (Lange et al. 2016). However, it is not necessary to go to this extreme for regional models to be more accurate. Much will depend on what the data (sample sizes) permit, but the analysts should consider what kinds of splits are pertinent and feasible. Is urban/rural a better split than models that distinguish by political/administrative unit (for example, state or department)? Does distinguishing between metropolises, mid-size cities, and smaller urban townships yield improvements in urban models? Do rural models improve if they are broken into major agro-ecological zones (for example, mountains versus plains, or desert or jungle versus moderate climates)?

A less common approach is to use multiple regional models that are fine-tuned for different program thresholds. If scores are being used to select households for multiple programs of different sizes, then a policy maker could use a single score for each household and a different eligibility threshold for each program based on program size, which might be based on the available budget or predetermined program objectives. Indeed, this

approach is currently taken in many countries using PMT. A policy maker could also choose to use a different model for each program on the basis that models can be optimized to target different parts of the income or consumption distribution and a set of program-specific models can be more accurate for each program than a single one-score-fits-all approach used for all programs. Nonetheless, using multiple scoring models requires considerably more time and effort to develop and can be difficult to communicate to policy makers and the public. Whether any improved accuracy warrants these complications is a trade-off to be assessed.

Choosing the Explanatory Variables

The choice of explanatory variables is key for modeling. It can be attractive to use many, even all the potentially suitable variables at once in the model without further processing, or deeper thinking on how the variables interact with one another. This can result in overfitting where the model is very good at predicting the survey data but not very good at predicting new data (which is how the model will be applied). Some practitioners use a stepwise approach that involves introducing (forward stepwise) or eliminating (backward stepwise) variables one at a time and using a statistical test to determine whether the model fit is improved.[43] This is a computationally efficient way of assessing model effectiveness, which improves upon assessing all possible variable combinations. However, a stepwise approach may not produce the best model. As James et al. (2013) note, if the best one-variable model uses variable A while the best two-variable model uses variables B and C, then the best two-variable model will not be assessed (variable A is selected from the first stage and only variable A + variable B and variable A + variable C will be assessed next). Alternative approaches from machine learning prevent overfitting due to the inclusion of too many variables. These approaches have started to be incorporated into traditional PMT development. A common method includes a penalty for complexity in the model.[44] This results in simpler models and tends to improve how well the models perform on new data. Common machine learning "penalized algorithms" include the least absolute shrinkage and selection operator (Lasso) and Ridge regressions. Ridge regularizes the model to prevent overfitting, while Lasso both regularizes the model and facilitates variable selection. Lasso has been used in several countries. For example, the recently developed PMT models in Iraq use Lasso,[45] as does the new poverty map being developed for Jordan.

Regardless of the process for incorporating variables in a model, various standard data analyses can be implemented. Before choosing the covariates, it is good practice to run basic data analysis to reproduce the official poverty and inequality statistics (which are usually constructed by the

national statistical office). This step is important as social programs are assessed on their capacity to mitigate poverty. Ensuring that the welfare metric is correctly built using the proper consumption or income aggregate, as well equivalence of scales and regional price adjustments, if any, is a precondition for having a good model. Poverty assessments and other such studies can be important sources to inform the analyst about the potential covariates that should be used in the statistical models for deriving a model.

Another good practice for selecting an initial set of explanatory variables is to start with traditional methods for exploratory data analysis among the variables that would be easily observable when an applicant is filling in program forms. A researcher must first "read" the data to understand its strengths and limitations and clean it to deal with missing observations and outliers. Using the frequency of responses (low-frequency responses can bias the predictors as they become noise in the model), sampling design, and sampling weights to reproduce core statistics made available by national statistical offices guarantees proper data manipulation later. Data visualization through scatterplots, histograms, box plots, normal plots, and the like can also be employed. Such analyses are at the core of any cause-and-effect analysis, and the approach should not be different for predictive modeling.[46]

Once the exploratory analysis is completed, further analysis that is more tailored to the traditional modeling is necessary. This stage comprises grouping variables into blocks and then analyzing each group separately, looking at variable correlations, including correlation with the dependent income/consumption variable. New variables can be created—such as "acceptable lighting material for the household"—based on the number of responses in different categories and their correlation with the dependent variable. For example, electricity access with own meter or community meter can be combined as *desired*; electricity without meter, oil, kerosene, or gas combined as *acceptable*; and candle and others combined as *unacceptable*.

In addition, interactions of location indicators with other variables are sometimes used to increase the predictive power of the model and lower exclusion and inclusion errors. The inclusion of location indicators may create separate thresholds and separate PMT weights for each location. The trade-off between capturing location-specific circumstances and maintaining a common threshold for all beneficiaries needs to be addressed explicitly as part of program policy. Nevertheless, the use of local-level indicators and estimation of different thresholds based on local poverty lines have provided better results in Honduras, Kenya, and Mexico, as well as the West Bank and Gaza, to name a few. One way to avoid this issue is to use real income or consumption measures in the regression, adjusting before modeling for differences in the cost of living across different locations. Then the location-specific circumstances

capture local aspects that relate to the real standard of living rather than differences in prices. Generally, this step should be done even when location-specific indicators are not being used.

In addition, the increasing amount of ancillary data coming from big data and the modernization of administrative data systems increase the ease with which local-level variables can be introduced in the model, which helps reduce model bias and variance (as residual location effects can greatly reduce the precision of the welfare estimates).[47] These variables are fixed at the enumeration area level; therefore, to incorporate them directly into the model, the data analyst may need to work closely with the national statistical office because enumeration area codes are often not available in the public version of the data made available to researchers. If the system that is used to code geographical areas in the ancillary data is different from the survey enumeration areas, a concordance between the two systems must be built, and the modeler will need to assess the trade-off between the required time and model improvements. In addition, machine learning can produce new variables from within the existing survey data, which can also improve accuracy; this is discussed in the following section.

The following list describes a nonexhaustive example of how to create groups:

- *Characteristics of the housing:* main source water, main source of cooking fuel, main source of lighting fuel, material of the walls, main toilet facility, material of the roof, number of rooms, room density, and expenditures on utilities[48]
- *Durables:* possession of satellite TV, vehicles, motorcycles, boats, and refrigerators
- *Land and livestock:* ownership of land, usage of land for agriculture, possession of livestock, and type of livestock
- *Characteristics of the household head:* gender, age, literacy, educational level, occupation, and disabilities
- *Characteristics of other household members:* share of adults working, average educational level, and number of adults with disabilities
- *Household size and type of family:* elderly living alone or couple, missing generation, nuclear family, number of children ages 0–5, number of children ages 6–14, number of youth ages 15–24, number of adults ages 25–59, and number of elderly ages 60+
- *Location and other local-level development variables:* urban, region, province, and enumeration-level aggregate information from other sources such as ancillary data to improve the precision of the measure of welfare.

It is recommended that for each group, researchers use the following steps to better understand the data and the problem they are trying to address:

- *Step 1: Assessing the dependent and covariate variables.* This step involves assessing proportions and means through descriptive statistics to identify the center, spread, and shape of each distribution. Frequency tables, bar charts, pie charts, and histograms allow identification of the mode (most common response) as well as categories with low response. In addition to simple descriptive statistics, some inferential methods are used to identify confidence intervals and perform significance tests.
- *Step 2: Assessing the correlation and associations of variables.* This step involves using descriptive statistics such as cross-tabulations to estimate conditional proportions, correlations, analysis of variance, simple regressions, contingency tables, paired differences, chi-square tables, nonparametric tests, and so forth, to understand the main correlations of variables to feed the predictive model.
- *Step 3: Assessing multiple correlations of variables.* This step involves assessing correlation in different ways. For multicollinearity we can use the variance inflation factor, which measures the correlation and strength of correlation between the explanatory variables in a regression model. To test the stability of the means and variances across variables, we can use the factorial analysis of variance. Where necessary, methods such as principal component analysis or other data compression methods can be employed to construct variables that reduce sampling variance.

Once the variables are assessed, the policy makers must move to the selection of models. In traditional PMT, applying steps 4 to 5 is recommended to select a final set of explanatory variables and model. For machine learning, the main step is to run the different models and let the computer select the best model (see chapter 8).

- *Step 4: Selecting the best model.* This step involves assessing the best approach for modeling based on the distribution of the dependent variable. If the dependent variable is a binary one, a logit/probit model is more appropriate, but for continuous variables, the researcher must choose between OLS, quantile regression, and the other models discussed earlier. Generally, when running PMT on a continuous dependent variable such as income or consumption, which are highly skewed, a logarithm transformation of the variable is used. The choice of the model involves running additional checks such as residual analysis to check the shape of the distribution and see how unusual observations affect the estimates. In addition, depending on the data analysis in the steps above, the researcher may decide to run different models for different geographical areas due to the representativeness and heterogeneity of the information per area. That is, different models per region, such as metropolitan, other urban, and rural areas, may be preferable to a unique model if the observable characteristics within each group are different—a fixed-effects

specification may not be the solution (see Elbers, Lanjouw, and Lanjouw 2003) to control for regional effects. However, the number of models is generally constrained by the representativeness of the survey sample and the trade-off between the time to construct multiple models and the improvement in predictive accuracy.[49]

Once step 4 is completed for each group, steps 3 and 4 can be repeated after grouping all the variables selected from each group. At this stage, new variables can be created by adding interactions to the model, while ancillary data at the lowest geographic level (such as enumeration areas), calculated from the census or obtained from ancillary data sources, can be added to the model specification to capture small area heterogeneity and improve prediction.

All the tests in steps 1 to 3 are important for measuring multicollinearity and instability of the coefficients caused by large variance. When high multicollinearity is present, the confidence intervals for the coefficients tend to be very wide and the t-statistics tend to be very small. The coefficients must be larger to be statistically significant; it will be harder to reject the null when multicollinearity is present. Detecting high multicollinearity is important and there are several warning signals. Most importantly, dropping variables should not generate large changes in a coefficient; model stability means that seemingly innocuous changes will not produce big shifts. Finally, the model selection implies that the set of explanatory variables in the group is composed of independent, unrelated groups, and the analysis of variance allows testing for a statistically significant difference between the groups, and how certain variables with large variance can bring noise to the estimates.

- *Step 5: Running a test for the null hypothesis that there is not specification error in the model selected and whether the residuals are homoscedastic, after estimations using the same model in step 4.* The first test is also known as an omitted variable test,[50] which tests the assumption that the error term and covariates are not correlated. The second test can be done using the Breusch-Pagan test to measure whether the residuals do not vary for lower or higher values of the covariates.

The final step, which is also applied for machine learning, is the simulation of the model performance for different cutoff points and special groups:

- *Step 6: Simulating coverage, performance, and predictive power for special groups.* Using the indicators presented in chapter 7 (coverage, distribution of beneficiaries, exclusion error or undercoverage, inclusion error or leakage, or benefit incidence) across the income or consumption distribution and for particular groups, such as urban/rural, female heads of households, and households with elderly living alone, the researcher can simulate the performance of the model against different thresholds (for

example, 2, 5, 10, 15, 20, and 30 percent) to estimate the model's predictive power.[51]

Once the modeling is complete, the scoring formula can be easily read by analysts and policy makers. Table 6.3 shows that predicted welfare increases for households whose heads have secondary or higher education and households with piped water inside the dwelling, electricity or solar panels for lighting, and gas/liquefied petroleum gas for cooking fuel.

Updating the Model and the Social Registry

Once it is calibrated, the model will continue to estimate the level of welfare with the same level of precision if the (partial) correlations between the variables selected in the scoring formula do not change, but of course they eventually will. Therefore, a good practice is to revisit the modeling approach, revisit the cutoff points, and retest the precision of the formula after two or three years, whenever a new representative household survey becomes available, or whenever the social registry or data system

Table 6.3   Illustration of PMT Weights for Selected Variables

		PMT weights
Caretaker characteristics		
Age	[ ] in years	0.0003
Educational level	0. No education	0
	1. Primary	0
Mark one of the codes below regarding your highest school grade attained	2. Secondary or higher	0.0881
What is household's main source water over the past month?		
Piped water inside dwelling	[ ]	0.4471
Rain, unprotected dug well/spring, river, lake, pond, or similar	[ ]	0
Other	[ ]	0
What is household's main source of lighting fuel?		
Firewood	[ ]	0
Electricity, solar	[ ]	0.1178
Other	[ ]	0
What is household's main source of cooking fuel?		
Firewood	[ ]	0
Gas/LPG	[ ]	0.5607
Other	[ ]	0

Source: Original compilation for this publication.
Note: LPG = liquefied petroleum gas; PMT = proxy means test.

recertification cycle is completed (in static systems), or whenever a major program expansion occurs. The following examples illustrate these cases:

- Georgia assesses the performance of its main antipoverty program, the Georgia Targeted Social Assistance Program (TSA), every year (thanks to the regular production of representative household surveys that include information on receipt of TSA). It updates the PMT scoring every four to five years (see box 6.6, which illustrates the type of analytical work and the political economy of updating the formula). The TSA program started in 2006, and the formula was updated in 2010, 2015, and 2020.

- In Colombia, revisions to the PMT formula are made with each new survey sweep that updates the SISBEN social registry.[52] With each new cycle, based on academic studies, the National Planning Department changes the modeling technique, the number and delineation of regional models, and the variables used. SISBEN I used one national model using principal components; SISBEN II used different OLS models for urban and rural areas; SISBEN III used three OLS models—one for the 14 cities, one for other urban areas, and one for rural areas; and SISBEN IV uses multiple machine learning–PMT models differentiated by department and rural/urban areas.

- The Palestinian Authority has reviewed its PMT formula as new data have become available. The Cash Transfer Program was created in 2010. The PMT used for its targeting was based on the then most recent available household survey, from 2007. The formula used regional cutoff points (one for Jerusalem, one for the West Bank, and one for Gaza). During 2011, the Palestine Expenditure and Consumption Survey 2009 was made available, and the formula was reassessed with the new data. The improvements in targeting accuracy that could be obtained by simply updating the formula were considered not enough to warrant a change; instead, attention was given to improving the delivery system.[53] A new round of the survey was fielded in 2017, data were made available in 2019, and since then, the Palestinian Authority has carried out the analytical work to update the formula, modify the regulatory framework, and adjust the delivery system. The Ministry of Social Development is now starting the process to recertify beneficiaries and apply the new formula as part of strengthening the Ministry of Social Development's social protection system.

- In Tanzania, revisions to the formula follow the availability of new data but with some lags related to when the data are available and when a new phase of program intake is due. The government piloted a community-based conditional cash transfer program during 2008–12,[54] which was subsequently scaled up into the Productive Safety Net Program. The process evaluation for the pilot suggested that the

implementation of the two-phase PMT it had used was problematic for logistical reasons. Thus, before the phase I scale-up in 2011, a new PMT model was estimated, still on the basis of the 2006 Household Budget Survey. The next round of major revisions to the formula was triggered by the next phase (II) of scale-up in 2015 and the availability of data from the 2011 Household Budget Survey. The formula was revised again in 2020 using the 2018 Household Budget Survey, again in time for a new phase of scale-up to full national coverage. All three rounds of revisions made in 2011, 2015, and 2020 included reweighting and adding/dropping a few variables.

As discussed in papers on poverty and inequality decomposition,[55] household welfare changes over time can be understood mostly in terms of changes in observed characteristics of the population and the returns to those characteristics. Changes in observed characteristics mean that the average characteristics of the poor have changed over time. For example, 10–15 years ago, possession of a cell phone or having connectivity to water or sanitation services could be associated with not being poor. Nowadays, cell phone coverage is extensive in most countries and large sanitation programs may have improved accessibility to potable water and sanitation in poor areas, so those variables are no longer close correlates of poverty in many countries (although a smartphone may be). Changes in returns to those characteristics mean that the value of a given characteristic has changed. For example, the value of a high school diploma relative to not completing high school may change as more people complete high school, or it may depend on the skills sought as technology changes or as trade and competitiveness policies, prices for major commodities, and the like influence the demand for different skills in the labor market.

Accounting for these potential determinants of changes in the distribution of welfare, periodic updating of a model should consider its various features, that is, whether (1) characteristics or returns to characteristics have changed, (2) adjustments are needed for use of the model by different programs, (3) new poverty lines/thresholds must be estimated, and (4) modeling errors due to past modeling limitations and low-quality data would improve if new data were used.

Poverty decomposition models such as the Oaxaca-Blinder decomposition can be used to assess the need for updating the traditional PMT model using the new data. The decomposition consists of estimating the difference in returns to characteristics over time by comparing the coefficients of the selected variables in the model for year *t* and year *s* and estimating the differences in the characteristics themselves by comparing the matrix of characteristics used in the model for both years.[56] The process is as follows:

(1) Estimate the same model used for estimating the PMT in year t on the new household data for year s to run a chi-squared statistic test to compare the equality of both vectors of PMT weights (price effect).

(2) Run a chi-squared statistic test to compare the equality of the matrixes of covariates used for the PMT model for years t and s (characteristic effect).

In practice, the assessment will generate two stylized situations:

(1) There is a price effect, for example, the vectors of PMT weights are different. The solution is to use the same PMT model but estimated on year s and adjusting the poverty line/threshold as the price effect implies economic growth and inflation or economic contraction with no changes in the determinants of poverty.

(2) There is a characteristic effect, for example, the distributions of the explanatory variables are different (or there can be both price and characteristic effects). The solution is to consider a general model change by adding or removing explanatory variables and adjusting the poverty line/threshold using year s data, as the determinants of poverty and the macroeconomic environment have changed.

Recent poverty assessments and other studies would indicate whether the poverty and inequality determinants may have changed. In this case, the test presented above will corroborate the findings of the poverty assessment, while the reports will guide the data analyst to identify new variables that could be added to the model.

Changing the model to add new variables has implications for the delivery system processes of social programs and the political economy. First, adding new variables may mean revising the full intake questionnaire, which involves changes in communication, outreach, software, training of staff, new intake and registration process, and so forth. Second, a new set of PMT weights means having new scores for each person and maybe new thresholds, so some current beneficiaries may lose benefits, while others may need to be included. Hence, there are winners and losers due to the new model, which requires new processes for onboarding, communications of rules and responsibilities, and so forth. Box 6.6 describes how this was handled in Georgia.

Another issue is that PMT models are based on household socio-economic characteristics, and transfers can have a direct impact on those characteristics. A poor beneficiary can use the transfers to improve their human capital and productivity, buy an asset, repair housing, and so forth. Such improvements caused by the transfers affect the characteristics that initially characterized the household as poor. Applying the updated PMT for a current beneficiary household would lead to a better score simply

BOX 6.6

Analytical Underpinnings and Political Economy of Updating the PMT Formula: The 2013–15 Reform of the Georgia Targeted Social Assistance Program

The Georgia Targeted Social Assistance Program (TSA) offers a useful illustration of the technical work and political economy of updating the scoring formula, the value added of regular reassessment of the scoring formula, and its implementation. The TSA was introduced in 2006, and its proxy means testing (PMT) was first updated in 2010. By 2013, several questions about the effectiveness of the program were debated by program staff, in policy circles or the media, which spiked before the national elections. A first concern was whether the formula was predicting the welfare of applicants with the same accuracy. There were concerns about weaknesses in design that could allow households to fool the system and about households concealing goods to gain eligibility for the TSA. It was believed that there were leakages to nonpoor families, and that beneficiaries were reducing their work effort when on benefits. The government decided to conduct a technical review of the program's effectiveness, including its scoring formula (Baum, Mshvidobadze, and Posadas 2016). Figure B6.6.1 presents the timeline of the program's key developments.

The objectives of the technical review were to validate and improve the effectiveness of the TSA: (1) to minimize inclusion and exclusion errors associated with the program, given the changing economy; (2) to remove from the PMT formula easily concealable durable goods, as there was a belief that households were indeed concealing them in an effort to be eligible for assistance; (3) to include new, easily verifiable, and potentially income-generating items[a]; and (4) to reduce the total number of variables used in the PMT formula. An impact evaluation was carried out to estimate whether the program generated work disincentives, and it proved the belief to be wrong (World Bank 2015). In addition, microsimulations using the most recent household survey data (2013) and the database of the social registry were performed to update the scoring formula (which includes two components, a consumption estimate and an estimate of the adult equivalents in a household, called a "needs index") and recalibrate the benefit level. This analysis focused on the winners and losers from the reform (due to changes in eligibility and the benefit level), disaggregated by area of residence and region, other household characteristics, and selected vulnerable groups (persons with disability, internally displaced persons, and single pensioners) (table B6.6.1). In terms of process, the

BOX 6.6 *(continued)*

Figure B6.6.1   Timeline of Key Developments in Georgia's Targeted Social Assistance Program

12/2014: Resolution No. 758: Methodology for assessment of the socioeconomic condition of socially vulnerable families

5/2015: Decree No. 215 of the government, which amends Decree No. 145

8/2015: Adoption of compensation measures

2–3/2014: Collection of additional module in IHS

7–11/2014: Revision of needs index

9–11/2014: Estimation of PMT based on new data

12/2014: Discussion of PMT models, introduction of new benefit structure for the TSA and CBP

12/2014–5/2015: Analysis of winners and losers

2–5/2015: Pretesting of the PMT with social workers

6/2015: Reassessment of TSA households begun

● **Technical work**   ◉ **Policy actions**

Source: Baum, Mshvidobadze, and Posadas 2016.

Note: The large hollow circles represent legislation (or equivalent), and the small solid circles represent technical work or other agreements. CBP = Child Benefit Program; IHS = Integrated Household Survey; PMT = proxy means test; TSA = targeted social assistance.

review recommended pretesting the new parameters before the national rollout.

The pretesting[b] was done between March and April 2015, once the law with the new eligibility formula was passed in December 2014. The pretesting confirmed all the predictions of the microsimulations and provided more precise estimates on winners and losers among small, vulnerable groups, which were oversampled for this purpose. For single

continued next page

BOX 6.6 *(continued)*

Table B6.6.1   Distribution of Winners and Losers (Intensive and Extensive Margin), by Demographic Group (percent)

Demographic categories	Always ineligible	Newly eligible	Newly ineligible	Benefit reduction	Benefit increase	Total
Residence type						
Urban	52.7	23.4	28.3	38.4	34.1	49.2
Rural	47.3	76.6	71.7	61.6	65.9	50.8
Female-headed household						
No	75.7	71.2	62.2	59.6	66.6	73.9
Yes	24.3	28.8	37.8	40.4	33.4	26.1
Number of children						
None	45.3	21.6	78.5	42.6	7.6	43.8
1 or 2	47.8	50.8	17.0	51.8	63.8	47.8
3 and more	6.9	27.6	4.5	5.6	28.6	8.5
Highest level of education for any member of household						
Lower secondary or less	1.9	10.2	7.7	8.2	2.0	2.7
Upper secondary	18.5	48.5	46.8	42.9	47.5	23.1
Secondary vocational	24.2	25.2	36.3	28.0	34.2	25.2
University	55.5	16.1	9.2	20.8	16.2	48.9

Source: Baum, Mshvidobadze, and Posadas 2016.

pensioners, persons with disability, and internally displaced persons, compensatory measures were designed, piloted, and implemented, which reduced the number of losers and increased the number of winners. By 2015, the government reformed the TSA to implement a simplified and more effective PMT formula. Compensation measures to reduce the losses of the losers from the reform were introduced in August 2015. Another round of revision of the scoring formula, focusing mostly on updating the coefficients and simplifying the formula, was implemented in 2020 (Honorati et al. 2020).

a. Most variables provided by households to the Social Services Agency are cross-verified against various databases from several sources, including the Ministry of the Interior (car registration), gas and electricity companies, the revenue service, and customs control.

b. The pretest sample comes from the Social Services Agency database of the TSA applicants that have applied for benefits since June 1, 2010, comprising 407,307 households. Using two-stage cluster sampling, 4,560 households were selected. Full interviews were conducted with 3,565 households.

because the household used the transfers to improve their living conditions, but that would not necessarily imply that they have higher autonomous incomes. One approach to get around this issue is to allow households to remain in the program for a time although their PMT score is above the entry threshold (for example, having different entry and exit thresholds). This is somewhat analogous to means tests with income disregards. In Mexico, for example, the exit threshold of the Oportunidades program was conditional on achieving concrete outcomes in food security, health, and education. For this reason, there was a higher exit criterion to accommodate improvements in beneficiary household well-being. There is a logic to this, but it implies treating households with the same PMT score differently, which contravenes the usual notion of horizontal equity. There is thus a delicate balance to be found.

Finally, new variables for use in both traditional PMT and machine learning can be identified and collected. Programs often have monitoring and evaluation systems that benefit from quantitative and qualitative assessments. Such evaluations can inform program administrators of relevant variables that are not currently used in the model or not available in national household surveys, which could improve prediction. Such findings must be discussed with national statistical offices to evaluate the possibility of adding such data in the upcoming surveys.

In summary, updating traditional PMT or a machine learning model implies not just running new regressions, but also thinking carefully about updating the full set of applicants and revisiting the overall implementation process. The timing of updating should consider the cycle of recertification for the social registry or the main programs that use the method to determine eligibility, and plan proper communications and a new strategy for excluding/including new households due to winners and losers caused by the changes. Moreover, improvements/investments in the interoperability of the information system may lead countries to move from PMT to HMT or means testing, as well as to define new strategies to reduce inclusion errors by applying other criteria as asset filters. Exclusion errors caused by flaws in the delivery system would remain unchanged (see chapter 3) with a model update, but they could be addressed in the associated round of data collection.

Data-Related Limitations and Considerations

Building PMT and machine learning models requires good data and good analysis. Many countries face one or more of three principal challenges on the data side: (1) limited periodicity of the surveys, (2) small sample sizes, and (3) sample design.

The data that are most commonly used to determine proxies and formulae are obtained from income and expenditure surveys collected by national statistical offices. These surveys are used for weighting the consumer price index and poverty and inequality studies, as well as for indicators of human capital and social development. In poorer countries, surveys are more likely to be fielded every five years or so; in upper-middle-income countries, statistical offices field them every two or three years. Having gaps of five or more years between surveys can affect the quality of PMT models as the predictors of poverty may have changed.

The addition of ancillary data can provide a good idea of the importance of certain kinds of local conditions, local infrastructure, and vulnerability to shocks as covariates to explain welfare. The small area estimation modeling of Elbers, Lanjouw, and Lanjouw (2003) shows that the best predictions for small areas come from having a different model for each area. However, most household surveys' sample sizes do not allow for such a fine breakdown. In traditional PMT, at least, researchers should allow for different slopes for different localized areas within regional or national models by using interactions. In addition, when small area models are not possible, an improvement that can be done is the addition of ancillary data from the small area into higher-level models.

Limited sample sizes and the sampling design[57] of household surveys can also affect the accuracy of the PMT formulae. Household surveys are designed to generate estimates at a particular stratum with enough observations to provide credible estimates with low standard errors, based on a representative number of clusters within a set of enumeration areas, and with a small number of observations per cluster (12 or 15 observations typically). Sampling weights (probabilistic weights) are given to each observation in each cluster and enumeration area for each stratum based on population and cluster characteristics, which can lead to different individual weights, to generate unbiased statistics at the stratum level. This approach is robust as it minimizes intracluster variance and maximizes between-cluster variance. It is efficient for most statistical tests and models because it directly addresses the bias-variance issues that haunt researchers and helps in understanding the average characteristics of the population. Nevertheless, for predictive modeling such as traditional PMT, the analysis relies on having greater variability to capture within-cluster differences and differences between populations; that is, predictive power is placed on the between-cluster variances.[58] For this reason, it is preferable to have (for a fixed sample size) a different partition that would guarantee that individuals within clusters are no more similar than individuals in different clusters. That is, it would be preferable to have fewer clusters and more observations per cluster to understand the within-cluster correlations for predicting the nature of their behaviors.

The sample size and the fact there is often only a single data set also have implications for the way the predictive power of a model is measured. Predictive modeling power is measured by its capacity to predict household welfare using a different set of data than that which generates the model. To do this, most frequently, researchers do data partitioning. Traditionally, this has meant splitting the sample into two groups, one for modeling (or training) and one for testing. This is an attempt to overcome different sources of prediction error: model variance and model bias.[59]

Some researchers use resampling methods, such as bootstrap or jack-knife resampling,[60] because more data are preferable for predictive modeling to have better control over bias and variance. More specifically, the approach of resampling allows better measurement of the variance, showing the variability of a model prediction for a given data point. An approach from machine learning that is becoming increasingly popular in PMT development is to split data into training and test data sets. However, instead of just splitting the data once, a standard machine learning approach (called *k*-fold cross validation) splits observations into multiple groups and takes turns estimating the models on some of the groups and testing them on other groups. For example, the data might be split into 10 random and equal groups (or folds). The model is estimated using nine of the groups as the training set and the last group as the test set. This process is then repeated after moving the test group into the training set and swapping one of the previous training groups out to be the new test set. Once all the groups have been in both the training and test sets, the model's performance is assessed by averaging the errors across all iterations. This has the advantage over just using a single split of ensuring that every observation can appear in both the training and test sets.[61]

Nevertheless, there is always a trade-off between bias, which is the error that measures the difference between the (average) prediction of the model and the correct value for a household and mainly caused by underfitting and overfitting, and variance. In predictive modeling, the aim is to have both low bias and low variance. When choosing a model and selecting variables, it must first be acknowledged that the predictions are mostly based on a single data set, meaning that there is greater control over bias and less control of variance without resampling. Moreover, Greene (2011) recommends the following: (1) acknowledging/respecting the estimator's properties; (2) addressing multicollinearity that can be identified when small changes in the data produce wide changes in the parameter estimates, or when coefficients have very high standard errors and low significance levels although they are jointly significant and the R^2 for the regression is quite high, and when coefficients may have the wrong sign or implausible magnitudes; (3) avoiding using pretest estimators that try to address multicollinearity by adding a third estimator (pretest), which is not

recommended as an ad hoc remedy for multicollinearity; and (4) understanding data measurement errors, which leads to a better understanding of data/modeling issues that have an influence on the bias-variance trade-off. This is important because at the core of this relationship, the analyst must deal with overfitting and underfitting of the model. As bias is reduced (better prediction of actual outcomes), variance is likely to increase (wider range of predictions in different data sets). In other words, in moving away from underfitting by adding more variables, the likelihood of variance sharply increases due to model complexity. As more and more parameters are added to a model, the complexity of the model rises, and variance becomes the primary concern while bias steadily falls. As Fortmann-Roe (2012) shows, there is a sweet spot, called model complexity, for any model where the level of complexity at which the increase in bias is equivalent to the reduction in variance, meaning that more complexity would overfit the model while less complexity would underfit the model. However, finding the right complexity level is not a simple task. There is no right or wrong way to find this balance, but as an acceptable prediction error can be set, the analyst can simply explore different levels of complexity and then choose the complexity level that minimizes the overall error.

When using statistical methods to create "pseudo" testing samples, resampling, or deciding to partition the current data set for testing and model calibration, researchers must always pay attention to and respect household sampling design. Household surveys include expansion factors/weights that indicate how many of the number of units in the population being surveyed each of the sampled units represents. Hence, using expansion factors/weights as sampling weights is necessary for the model estimation of proportion, means, and regression parameters. When partitioning or resampling to create pseudo test sampling, researchers must first ensure that the sampling design is incorporated into the process. Selecting a process by convenience implies that the results may not be representative of the population, adding bias in the estimates generated. Moreover, survey balance can disappear without proper definitions of strata, clusters, and enumeration areas. Respecting sampling design and adding a reweighting process to correct the sampling weights is needed to guarantee that both the treatment and testing (holdout) sample, as well as all pseudo-samples generated by statistical methods, still represent the population for which the household survey was designed.

Hence, testing whether the household survey data set of the potential regressors/covariates distribution matches that of the population of interest is key before running predictive models. Comparing the descriptive statistics and covariate distributions (of both the training and holdout or pseudo-samples) with other data sources, such as censuses, Demographic and Health Surveys, labor force surveys, and so forth, can indicate how

accurate PMT prediction might be, as well as identify some variables that would need to be in the model regardless of their significance level.

An intriguing way of overcoming some of the limitations of traditional household surveys that are designed to be representative of the general population is to use data that are more keyed to the poorer population. Some surveys oversample to be able to estimate proper indicators among the poor. For example, in Mexico, the Socioeconomic Conditions Module was created as a complement to the National Survey of Household Income and Expenditure (ENIGH), with the purpose of providing the necessary information for the National Council for the Evaluation of Social Development Policy to carry out the measurement of multidimensional poverty at the national and state scales. The ENIGH and Socioeconomic Conditions Module combination was implemented starting in 2008, and to be able to estimate poverty and multidimensional poverty by state, the ENIGH sample, which was 35,146 households, almost doubled to reach 70,106 households. Another option is to use data or samples from the social registry, which by design is concentrated among the poorer (see box 6.7).

BOX 6.7

Using Data from the Social Registry to Update PMT Formulae

As more countries are investing in developing integrated social registries to support the delivery of one or many programs, the data gathered can be used to calibrate beneficiary selection formulae in ways that overcome some of the limitations of traditional household surveys. Paes de Barros et al. (2016) show that basing a proxy means testing (PMT) formula on the Brazilian Unified Registry of Social Programs (Cadastro Único, or CadÚnico) would improve eligibility determination for the Renda Melhor guaranteed minimum income–type program created by the state of Rio de Janeiro in 2011, to complement the Bolsa Familia program for extremely poor families.

Until 2016, the Renda Melhor program offered each of the Bolsa Familia beneficiaries in the state of Rio de Janeiro an extra benefit that would complement the post–Bolsa Familia beneficiary income by the income gap needed to reach to the extreme poverty line of R$100 (US$29) per capita. Therefore, to be in the program, a family should be eligible for the Bolsa Familia program through its means-test approach

continued next page

BOX 6.7 *(continued)*

from the CadÚnico, which is used by more than 20 federal social pro-
grams and by many states and municipalities for developing social
policies and programs at the local level.[a] However, while Bolsa Família
has been using only declared income as the eligibility criterion, the
Renda Melhor program has sought to make full use of the variety of
information contained in the CadÚnico. Policy makers started using a
secondary approach to estimate the full income of households eligible
for Bolsa Família under the assumption that the declared income was
not a good representation of the household's permanent income.
Hence, they used a PMT to estimate this permanent income based on
the information available in the national household survey data and the
CadÚnico. This approach seems promising under the conditions of
having a social registry that is large enough, regularly updated,
dynamic, and assessed as providing a good representation of the poor
as these data can work almost as a census of the poor population. On
such basis, the authors conducted the following experiment:

- From the CadÚnico, a probabilistic two-stage sample of 4,000
  households was extracted and guaranteed representativeness of
  the CadÚnico population in the State of Rio de Janeiro, and for
  three income groups: (1) families with self-declared per capita
  income below half the minimum wage,[b] (2) families living in house-
  holds with total self-declared monthly income of up to three times
  the minimum wage, and (3) families with self-declared incomes
  greater than three times the minimum wage and receiving any
  social program from the three levels of government (federal, state,
  and municipal).
- A special survey was administered to households in the sample.
  The survey questionnaire was designed to mimic the CadÚnico
  form, with a few extra variables such as a more detailed income
  module to address the specific needs of the assessment.
- Once the 4,000-household data collection was completed,
  researchers merged the survey and CadÚnico data to have two
  measures of income: (1) the self-declared income from CadÚnico
  and (2) the full and detailed income collected through the survey.
- Researchers estimated the underreporting elasticity of the
  CadÚnico income by comparing the two incomes and
  reestimated the PMT model used by the Renda Melhor program.
- The new PMT predictor was then applied to all the data for the
  State of the Rio de Janeiro to determine eligibility for the Renda
  Melhor program.

BOX 6.7 *(continued)*

- The researchers estimated that a new PMT could reduce the Renda Melhor program's inclusion errors from the original PMT model by 7–33 percent. Exclusion errors remained at the same level.

This exercise highlights the importance of continuously using the information in the social registry as it contains a denser concentration of the poor and vulnerable population than in any household survey. The authors believe that eligibility determination is strengthened by having high-quality information gathering at the application stage, requiring consistent definitions and concepts that are consistent with the national statistical office questionnaire for collecting income/consumption information, and allowing program administrators to define a recurrent monitoring and evaluation strategy to measure the quality of information.

A proposed approach would be routinely sampling a subset of applicants and gathering from them a longer questionnaire as in national survey samples. The sample would also allow administrators to have a deeper understanding of population and poverty dynamics through using statistical pattern recognition, which could in turn help to calibrate circumstances that would trigger the need to update information, or for provision of additional documents for checks and home visit inspections. However, to avoid underdeclaration of income in these survey efforts, it is important to ensure that the process for collecting such data is not seen as a condition for eligibility decisions but as regular monitoring and evaluation of government strategies. For example, data for this sample can be collected later and not at the stage of enrollment in the program.

a. Cadastro Único (CadÚnico) was officially created in 2001 through a Presidential Decree (#3,887) by Fernando Henrique Cardoso. The implementation of the first large-scale expansion of the CadÚnico started in 2003 during the phase of consolidation (2003–05) of four cash transfers schemes (Bolsa Escola, Bolsa Alimentação, Cartão Alimentação, and Vale Gás) into the Bolsa Família program that formed the initial largest base for the CadÚnico. As the CadÚnico matured during 2006–09, it became the gateway for benefiting from low-income families social policies. Between 2010 and 2013, CadÚnico version 7 introduced online synchronization with the federal center and other systems, such as pension systems. The CadÚnico management involves the three levels of the federation: municipalities, states, and the federal government. The municipalities are the main actors as they are in charge of its implementation (for example, they identify low-income families; interview, collect, and register data in the national database; keep data updated; promote continuous capacity building to agents; and provide and maintain adequate infrastructure in the centers; keep and protect confidential information; and take measures to control and prevent fraud or registry inconsistencies). The states have a more planning and capacity-building role to provide municipalities the right skills and develop specific actions to register

BOX 6.7 *(continued)*

traditional and specific populations, such as quilombolas, indigenous, and home-less populations. The federal-level managers support coordination, supervision, and monitoring; define strategies and instructions to improve the quality of the information registered; and give financial support to municipalities and states to strengthen their capacity to manage and run the CadÚnico. At present, more than 20 federal social programs make use of the CadÚnico, which is also used by many states and municipalities for developing social policies and programs at the local level. Today, CadÚnico contains the details of 28 million families, of which 14 million are beneficiaries of the Bolsa Familia program.

b. In 2019, the minimum wage in Brazil was set at R$998, equivalent to approxi-mately US$249.27 a month or US$8.40 a day, as of October 2019.

PMT's Predictive Power beyond Chronic Poverty

PMT models are usually not good at predicting vulnerability to poverty from shocks. The chronic poor are often more vulnerable to falling into poverty from shocks, as they frequently live in more hazardous places, such as along railway tracks or places that are more vulnerable to climate change, and work in more occupations with significantly variable incomes. They also have less ability to cope with shocks when they happen. For these reasons, chapter 3 stresses that getting support out quickly through existing social assistance programs when a shock first happens is a good strategy even when the authorities have not yet identified the most affected people. However, being able to predict vulnerability to shocks before they happen can help guide programs such as insurance or incentives to use insurance or other risk-management strategies.

PMT models are not designed for identifying households after they suffer a shock. These models are designed to identify chronic poverty and low incomes based on proxies that are fixed or change only slowly over time, such as housing quality and demographics. As a consequence, when a household suffers a shock, whether idiosyncratic or covariate, its PMT score may not change or change only a little. For example, the household composition will likely stay the same, or the likely changes, such as sending a child to live with relatives, will make the score higher. Assets accumulated during better times will remain in the household unless they are sold to cope with the shock. Housing quality will not change unless the household moves to cheaper housing to cope with the shock.

Other models have been shown to perform better than PMT at identifying transient food insecurity. Schnitzer (2019) studied the precision and relative performance of PMT and Household Economic Analysis (HEA) for identifying the poor and vulnerable in Niger. Her findings show that given

the strong correlation between location and food insecurity, geographical targeting could be especially effective if the aim is to respond to food crises. Further, PMT performed more effectively in selecting persistently poor households, while HEA performed better in selecting transient food insecure households. Specifically, PMT resulted in lower inclusion errors by 29 percentage points than HEA based on persistent poverty rates, but higher inclusion errors by 18 percentage points than HEA based on transient food insecurity. Therefore, to address vulnerability, some update/adjustment of the model may be considered.

PMT models can be adjusted to account for different sources of vulnerability: poverty induced and risk induced. Skoufias and Báez (2021) distinguish between poverty-induced vulnerability when the predicted household consumption falls below the poverty line even without shocks, due to a lack of physical and human capital assets, and risk-induced vulnerability when predicted household consumption would be above the poverty line if not for a shock.

A multilevel model can show this decomposition as well as the extent to which vulnerability to poverty is associated with idiosyncratic and covariate shocks. Using only cross-sectional data, Skoufias and Báez (2021) demonstrate a PMT model that summarizes both idiosyncratic and covariate shocks. Figure 6.3 shows the vulnerability rates from the different sources

Figure 6.3   Poverty-Induced and Risk-Induced Vulnerability in Five African Countries
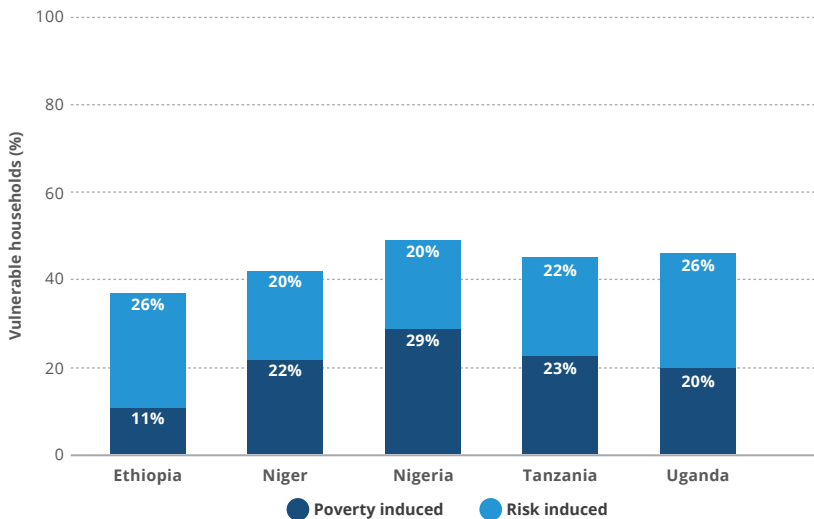
Source: Skoufias and Báez 2021.

for five African countries, decomposed into poverty-induced and risk-induced sources; the relative importance of each varies by country. Figure 6.4 shows the relative contributions of idiosyncratic and covariate-based risks to risk-induced vulnerability in Ethiopia where it can be seen that in both waves, the contribution of covariate shocks to vulnerability is higher in rural areas relative to urban areas.

A PMT model that accounts for ex ante vulnerability can also account for different exposures to shocks. Climate-related and other shocks lead to different potential states of the world. A household's poverty status is the realization of its consumption in the state that comes to pass. If the realized consumption is below the poverty line, the household is classified as poor; poverty status is backward looking (or ex post). Vulnerability is forward looking (or ex ante). In figure 6.5, the range of a household's outcomes under different future states of the world is presented. Some households (A and I) will be poor in all states of the world, while some will never be poor (B, E, and H). However, for the remainder, their realized poverty will depend on which state of the world comes to pass; their vulnerability to poverty depends on the risk of natural disasters and other shocks. Historical data on localized natural disasters and drought combined with realized household poverty outcomes can be used to predict which households will be at risk in the future. Such models can be used to prioritize the poor or vulnerable for covariate risk-mitigating social protection programs or public insurance schemes, helping administrators to manage covariate shocks, as most of the programs are designed to address idiosyncratic risks.

Figure 6.4   Decomposition of Vulnerability in Ethiopia
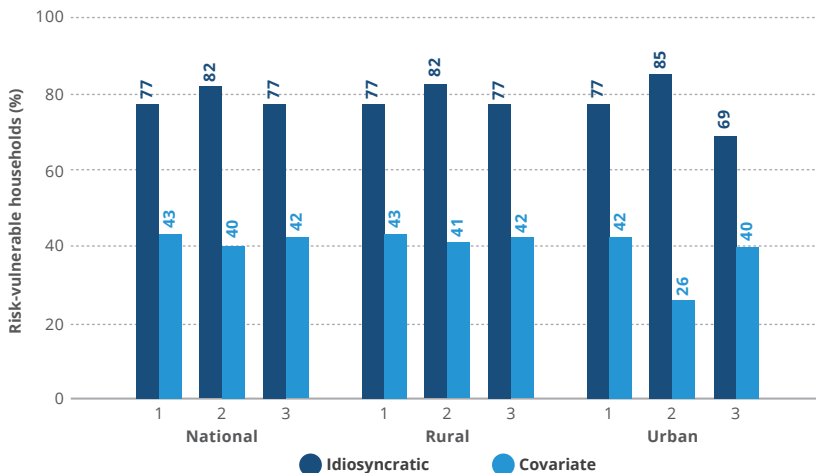
Source: Skoufias and Báez 2021.
Note: 1 = 2011–12; 2 = 2013–14; 3 = 2015–16.

**Figure 6.5   Vulnerability to Poverty Characterized by the Mean and
Variance of Welfare**
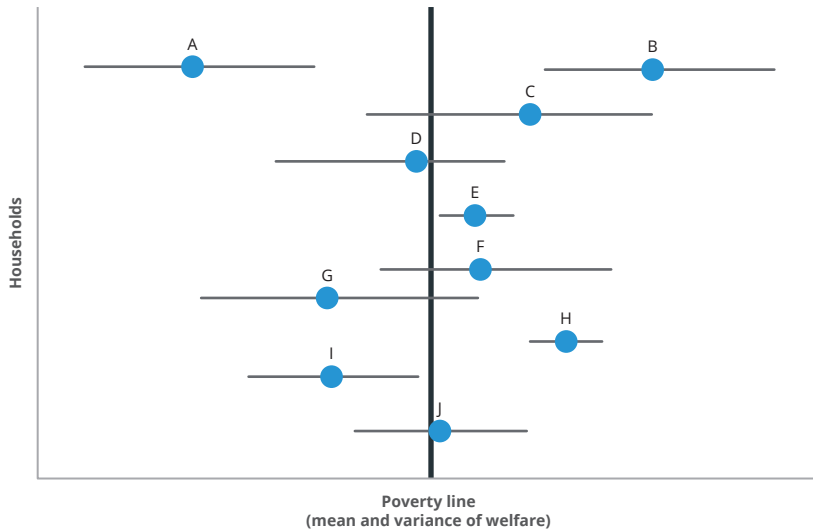
Source: Skoufias et al. 2019a.

To address vulnerability, the PMTplus model is presented by del Ninno and Mills (2015) and Leite (2014). PMTplus expands a traditional PMT model to incorporate ancillary data at the lowest administrative level possible, measuring shocks or vulnerability, to adjust for the impact of those shocks. Therefore, PMTplus is a variation of PMT that incorporates the impact that a major shock (drought, flood, incapacitation, or death of an adult family member) may have on households. Panel data (observations of the same household over multiple time periods) are a first-best option for this type of measurement. However, as panel data are not generally available, the PMTplus technique is appropriate for cross-sectional (single observation at one point in time) data sets (Skoufias and Báez (2021) call this the simulation approach).[62]

PMTplus is an extended model with fixed-effect variables that represents the impact on a local area exposed to a shock. Household data on shocks, georeferenced climate data, and community data can all be used to identify exposure. del Ninno and Mills (2015) show that aggregate climatic shocks can be estimated using widely available and detailed georeferenced information on historical rainfall from the US National Aeronautics and Space Administration's Langley Research Center POWER project website. Variations in rainfall from historical trends can be employed to obtain more

nuanced estimates of climatic impacts on PMT scores. The advantages of this approach are that data on aggregate shocks are often readily available and the estimation methods are the same as those used in the PMT. The disadvantage is that the use of aggregate information on covariate shocks is, essentially, a form of geographic targeting but the same impact is imputed to each household. Microclimates, geography, soil conditions, as well as farm practices may expose to drought or flooding only a portion of households in the same aggregate climatic conditions, yet all households in that location will be modeled as affected.

Leite (2015) simulates the impact of shocks in Kenya using the PMTplus approach, showing that small, shock-related adjustments to the PMT to reduce inclusion errors in times of shocks are possible. Simply applying the estimated impact of the shock on welfare to correct the cutoff point for the PMT and using an indicator of food insecurity, such as the World Food Programme's Food Consumption Score, makes it possible to identify households that are vulnerable to poverty. Leite (2015) suggests that the best way to increase the precision of the selection would be to examine the most up-to-date geographic data to identify the shock-affected areas (geographic targeting) and then carry out a quick data collection exercise to gather food insecurity indicators to improve the precision of the model.

PMTplus has been applied to several African countries to shed light on vulnerability rates and sources within a traditional PMT model. Skoufias et al. (2019b) explore some of Hill and Porter's (2017) ideas and use a hierarchical model based on Günther and Harttgen (2009); the analytic framework adopted is complementary to del Ninno and Mills' (2015) PMTplus approach as well as the Listahanan PMT.[63] Skoufias et al. aim to predict households' ex ante vulnerability to typhoons in the Philippines (Skoufias et al. 2019a) and variations in rainfall and the incidence of drought in Ethiopia, Niger, Nigeria, Tanzania, and Uganda (Skoufias et al. 2019b). They find that the exposure of a *barangay* (village) in the Philippines to a typhoon in the six months before the month of the interview is associated with a statistically significant decline in total per capita expenditures and specifically food and protein per capita expenditures. The estimated coefficients from the regression model were also used to estimate ex ante household vulnerability to poverty (the likelihood of household consumption falling below the poverty line) in the event of future natural disasters of different intensities.

PMT, Machine Learning, and Big Data: What Do We Know?

With more computational power, there is a new branch of research to investigate the promise of using more sophisticated algorithms to improve PMTs.

The traditional PMT models mentioned above historically used parametric regressions (OLS, logit, and quantile regressions). Machine learning algorithms use both parametric and nonparametric models that are more computationally intense (for example, nonlinear models and tree-based models) to score household welfare based on various observable household and location variables. The main traditional PMT models were briefly summarized in the section of this chapter on PMT; this section highlights the various new machine learning models along with what is known about how effective they can be; chapter 8 provides a fuller and more technical presentation of machine learning results by Areias et al. (forthcoming).

In addition, the new trends in big data (for example, remote sensing satellite data and CDRs) offer new proxies on which PMT models can be built. Machine learning can be applied to traditional survey data. It can also be applied to various forms of new big data. Earlier, this chapter described how big data, such as remote sensing satellite data, social media metadata, and CDRs, can be used to make geographic targeting much more accessible and frequent. In addition to reviewing the new machine learning models, this section considers the use of big data in PMT models, as well as the aforementioned ancillary data, to select households rather than small area locations.

Improving PMT with Machine Learning Models

The main difference between the traditional regression models usually used in PMT and the new machine learning models is their approach to model variance and model bias. Model variance is the degree to which a model would change if it had been trained on different data, and it is hoped that the model estimates do not vary too much across different training sets since high variance leads to lower precision. Model bias is the error that comes from errors in the model specification that uses simple linear relationships when most of the relationships are not linear. Traditional regression-based PMT focuses more on finding unbiased models, which does not mean that they have the lowest errors for prediction as they still can have high variance. Machine learning approaches accept a trade-off by introducing some error due to model bias in exchange for reducing error due to model variance in the estimation, thereby focusing on out-of-sample accuracy rather than correct model specification. The models are successful if the decrease in the error due to reduced variance is significantly more than the increase in error due to introducing bias. Moreover, the fact that machine learning methods are concerned primarily and explicitly with controlling model variance has allowed them to use very high-dimensional data, such as CDRs and satellite images, as useful inputs for estimation problems, without risking model generalizability.

There are various approaches in the machine learning literature. This section highlights four categories of models that are commonly used: robust models, penalized regression models, nonlinear models, and tree-based models, which use different optimization algorithms and can be contrasted with traditional PMT models. A brief and high-level overview of the differences in the modeling philosophies—and not an explanation of the differences between individual algorithms (model approach for maximization of the functions)—for each of the models is presented in box 6.8. A fuller treatment is provided in Areias et al. (forthcoming), based on James et al. (2013) and Kuhn and Johnson (2018).

BOX 6.8

Machine Learning Models That Are Commonly Used

Robust Models

Linear models (ordinary least squares [OLS] regressions commonly used in proxy means testing) can be overly influenced by outliers. If these outliers violate the assumptions on which linear models are based, the resulting performance of the linear model can be poor. Robust models are "robust" (less sensitive) to outliers. While they are computationally more intensive, robust models have less restrictive assumptions and thus can perform better across a wider range of data.[a] The main algorithms are robust linear regression and various quantile regressions with different methods for variable and quantile selection.

Penalized Regressions

Penalized regression models use shrinkage methods, in that they shrink the OLS coefficient estimates toward zero by introducing a penalty term on the coefficients. The objective of shrinkage methods is to reduce the variance of the models significantly with only a small increase in bias, thus reducing overall model error. This can be the case particularly when multicollinearity between explanatory variables is high. The basic approach is the same as OLS but introduces a penalty on the size of the coefficient (the degree to which a variable explains the outcome of interest). If a variable does not significantly improve the model, its role in the model is reduced. The main algorithms are Lasso regression, Ridge regression, and elastic net regression.

continued next page

BOX 6.8 *(continued)*

Nonlinear Models

Most of the models that are used for targeting applications are linear models. These models can largely be adapted to nonlinear relationships in the data—for example, by adding polynomial terms (age and age squared)—but this requires knowing in advance the nature of the nonlinearities. Among the range of nonlinear models, each works in different ways, but they are related in that the analyst does not need to know the nature of the nonlinearities in the data ex ante. The main algorithms are multivariate adaptive regression splines; *k*-nearest neighbors, computational neural network, and support vector machines.

Tree-Based Models

Tree-based models are a subclass of nonlinear models that differ from the other categories of models considered so far in that they use if-then statements to partition the data, rather than a regression. At their most basic, tree-based models split the data into two partitions, based on whether each observation's value on an explanatory variable is above or below a certain cut point. For example, is the household size more or less than four? Within each split, a further split is made based on the observation's value on a second explanatory variable and a second accompanying cut point. This continues until each branch of the tree reaches a terminal node, upon which the remaining observations at that node receive a predicted outcome, which could be as simple as a fixed value or the mean value of outcomes at this node from the training data, up to a regression-based prediction using all the explanatory variables. The main algorithms are all ensemble approaches[b]: random forest, random forest with quantile loss, gradient boosted regression trees, and gradient boosted quantile regression trees.

a. Areias et al. (forthcoming) look at five robust models: (1) robust linear regression with principal components, (2) quantile regression, (3) quantile regression with cross-validated quantile, (4) quantile regression with Akaike Information Criterion variable selection, and (5) quantile regression with Akaike Information Criterion variable selection and cross-validated quantile.

b. Tree-based models are popular for three main reasons: (1) they generate conditions that are highly interpretable and easy to implement; (2) they do not require specifying the relationship between the explanatory and outcome variables ahead of modeling; and (3) they handle missing data and implicitly conduct variable selection. At the same time, two well-known weaknesses are (1) model instability, and (2) predictive performance that can be beaten by other approaches. Ensemble methods have been developed to address these issues. This chapter looks at both basic regression trees and ensemble approaches, which build on them.

Kuhn and Johnson (2018) summarize the different models and algorithms and their characteristics along various dimensions. As is further summarized by Areias et al. (forthcoming), "there is no one ML [machine learning] model which is uniformly better than the others; the applicability of a technique is dependent on the type of data being analyzed, the needs of the analyst and the context of how the model will be used." Particularly of note is that the traditional PMT models (linear and logistic regressions) are easily interpretable and computationally easy with no tuning parameters, but they require significant preprocessing, have no automatic variable selection, and are not robust to predictor noise.

Does Machine Learning Improve Prediction Accuracy over Traditional PMT Regressions?

A few studies compare machine learning models relative to traditional regressions when using standard household survey data. The Areias et al. (forthcoming) exercise uses data from 17 Sub-Saharan African harmonized household surveys (100 subsamples each) to test 19 algorithms. It evaluates them on four metrics, which is the most systematic assessment to date. Shrestha (2020) uses the Mongolia Household Socio-Economic Survey 2018 to compare the current PMT model prediction with seven algorithms,[64] using training data—a 70 percent random subset of the Household Socio-Economic Survey 2018 data—and compares their performance in the test data not used for estimation of the remaining 30 percent. McBride and Nichols (2016) run a similar exercise with national household survey data from Bolivia, Malawi, and Timor-Leste, using the United States Agency for International Development poverty assessment tool and base data to compare the gains in accuracy of machine learning against PMT. Ohlenburg (2020b) compares seven machine learning models[65] with OLS and stepwise OLS for Indonesia.[66]

The first result is that an algorithm's performance depends on: (1) the targeting metric it is optimizing and by which it is assessed, and (2) the type of targeting problem it is asked to solve. Areias et al. (forthcoming) show that algorithms perform differently on different metrics. For example, consider an algorithm in the robust class.[67] This algorithm performs poorly on a standard machine learning metric (root mean square error) but is one of the best relative performers on a targeting measure that favors lower exclusion error (F_2).[68] At the same time, it performs poorly on another targeting measure that favors a balanced lower inclusion-exclusion error (Matthews correlation coefficient [MCC]; see chapter 7). This suggests that an algorithm's performance depends on the policy maker's objectives; if she wanted to reduce inclusion and exclusion errors in equal measure, she would not choose this robust measure. Conversely, if she cared most about

exclusion error, she might consider it. Similarly, the decision between using household scores with a poverty line versus a poverty quota approach for eligibility has implications for the choice of algorithm. The same robust algorithm is one of the best performers on F_1 and F_2 measures when the line approach is used, but it is one of the worst when the quota approach is used. Some of the other models reviewed showed similar swings in results. McBride and Nichols (2016) also find that the differences between machine learning and traditional approaches depend on the metric being used; on one metric, machine learning approaches are better and on another, they are worse.

Moreover, this small literature[69] finds that the gains in precision compared with traditional PMT regressions are marginal at best. All three studies show that even when there are clearly preferred algorithms on statistical grounds, generally the differences in performance are not important in magnitude. Areias et al. (forthcoming) find that most of the differences in model performance are less than 5 percent. They also highlight that in their study, the two best performers across the line and quota approaches are a machine learning model and a traditional regression,[70] while another machine learning model is clearly inferior across all performance measures and modes of implementation,[71] although there are nonperformance-based considerations as well.[72] Shrestha (2020) concludes that nonlinear machine-learning algorithms do not necessarily help improve targeting performance compared with a simple PMT recalibration in his study using Mongolia's most recent household survey data. Machine learning models failed to improve PMT performance in all sublocations of the country with different algorithms except one where the improvement was on the order of 2 percent. Moreover, the analysis shows that no single algorithm consistently outperformed the other ones. Ohlenburg (2020b) notes that in Indonesia, although some of the machine learning models perform better than the traditional PMT, their gain over a traditional regression is surprisingly limited considering that PMT should play to the strengths of machine learning methods.

Although the results are from a limited range of countries so far, they provide important considerations for the use of machine learning estimation in other settings. As Areias et al. (forthcoming) indicate, performance can vary across different dimensions, including the choice of targeting metric and the policy maker's objective, program size, and scoring implementation approach. Does a policy maker prioritize reducing exclusion error or more balanced inclusion and exclusion errors? Is she trying to use the same score to determine eligibility for programs of very different sizes? Is a threshold or quota approach being used?[73] Regardless of whether the precise results of this work extend to other settings, they indicate that there is not necessarily a universally best algorithm and that if an optimal algorithm does exist for a particular context, context will matter. Whether an

exhaustive search for the optimal algorithm is justified by the often limited real-world (if not statistical) differences between model outcomes is an important consideration for policy makers.

The ability to optimize machine learning algorithms for different measures of performance means that one approach may be one of the best performers on one metric and one of the worst performers on another metric, making it more important for policy makers to specify their preferred measures. Policy makers may also have different sensitivities to different targeting errors depending on the program and country. Some may be most worried about exclusion errors, prioritizing minimizing the number of poor who are mistakenly excluded from the program at the expense of including more nonpoor than they might under an alternative scoring approach. Others may be more worried about inclusion errors, as high-profile mistakes, such as including the mayor's spouse or a well-known businessperson, undermine the credibility of the targeting system. Yet others may prefer a balanced approach to minimizing inclusion and exclusion errors. Areias et al. (forthcoming) provide a few examples of machine learning models that are among the most accurate on one performance measure (such as errors of inclusion or exclusion) and one of the least accurate on another. This may not be a common occurrence, but it underscores the need for comprehensive evaluation of all models against policy makers' objectives and intended uses.

Transparency is a concern with PMT, but it is accentuated with machine learning models. Even a single traditional PMT model is often subject to the criticism that it is a "black box," both because the statistical approach can be difficult for a general audience to understand and because the exact scoring formula is often kept confidential to avoid households gaming the scoring system. Such criticisms could become louder when that black box is not a simple OLS or related regression but a neural network with scoring criteria that are very difficult to interpret and even more difficult to explain how they were derived; often the modelers themselves do not know how the algorithm got to its final destination. In contrast, the key advantage of the traditional PMT models is that a good model allows the researcher to read the coefficients and tell a story about who would be the beneficiaries of the program.

However, there are limitations to studies that only apply machine learning models to standard survey data. The systematic assessment of models in Areias et al. (forthcoming) is only a static assessment based on static surveys. It applies different PMT models and machine learning algorithms to the standard household cross-sectional survey data used to predict whether a household is below an income or consumption threshold. The conclusion is that machine learning does not offer significant improvements against traditional PMT with existing data; it says nothing about machine learning

improvements with new data. McBride and Nichols (2016) indicate that the conservative gains in accuracy from machine learning methods seem to be due to data limitations. Ohlenburg (2020b, 25) notes that a key limitation of his Indonesian study is that "the relative similarity of results between OLS-based approaches and various machine learning methods might imply that lack of additional information in the data is a greater bottleneck than variable or method selection." He also notes that no attempt was made to do feature engineering, which is discussed next.

Machine learning models could benefit from the dynamic data collection of the social registry (box 6.7). Given the data-hungry nature of machine learning approaches, they may offer significantly more promise than traditional regression-based models. For example, as the amount of information on households in a country's social registry expands, the information on each household changes over time. As more households enter the registry, machine learning algorithms, particularly unsupervised and deep learning approaches, may be able to improve performance by incorporating the new information. In other words, it is assumed that machine learning systems improve as more data are available or, according to Ohlenburg (2020a, 10), "as they learn, much like people, they should also keep learning as they keep performing a task, as people do."

Noriega-Campero et al. (2020) go beyond the earlier studies to incorporate verified income data from the social registry and use feature engineering to develop new variables from the existing survey data. The paper documents the development of machine learning–based PMT models for use in the social registries in Costa Rica and Colombia; they were adopted for implementation in the former and not in the latter. In particular, this real-world example improves on earlier work by making more data available to the models. First, the study pools multiple years of surveys to increase the number of observations, as was done in Indonesia to facilitate district-level models, providing samples of 22,000 in Costa Rica and 462,000 in Colombia. Such an approach would also benefit traditional regression-based PMT models but should only be used in countries where surveys are conducted close together so that the relationship between income/consumption and proxies is relatively stable. In many developing countries, surveys are five years or more apart; thus, pooling is not advised in such cases (this approach was not possible in Areias et al. [forthcoming]). Second, the study matches the survey data to administrative data on verified household income (in a sense, making this closer to HMT than PMT), which leads to a significant improvement in both traditional and machine learning models.[74] Finally, the study derives new variables from the existing survey data (as selected by experts):

> Second, statistical features, including means, modes and entropies for all individual-level variables of household members, such as age, gender, and

> education. Lastly, deep features, generated by a recursive neural network that condenses information of the individual-level features into a one-dimensional encoding—a technique akin to the AI subfield of multiple instance learning (MIL). (Noriega-Campero et al. 2020, 243)

Noriega-Campero et al. (2020) find significant improvements in prediction accuracy, which are largely derived from incorporating new data and variables; traditional models also show considerable improvements with the new data. Noriega-Campero et al. simulate targeted programs with coverage equal to the national poverty rate (22 percent in Costa Rica, 28 percent in Colombia).[75] The simulations indicate improvements of 20 and 26 percent for Costa Rica and Colombia, respectively, when comparing a traditional regression model (quantile linear regression) with standard survey variables to the best performing machine learning model (gradient boosting) with statistical and deep features. As figure 6.6 indicates, it is the inclusion of the statistical and deep features that drives the majority of the improvement. Gradient boosting with expert (standard) variables shows only a modest reduction in errors compared with quantile linear regression with expert variables. Quantile linear regression with the new variables outperforms gradient boosting with the standard variables and can provide over half of the improvement that gradient boosting with the new variables can.

Better survey and administrative data are likely more important than machine learning–driven feature generation and estimation in improving PMT accuracy. Despite the results of the paper, Noriega-Campero and his associated organization, PROSPERIA (https://www.prosperia.ai), consider feature engineering and algorithm choice to be less important in improving PMT prediction than having updated survey data and a wider range of quality variables. They suggest a ranking of factors, from most to least important, that would improve PMT (PROSPERIA 2021):

1. Updated survey data
2. Quality and quantity of usable variables (whether from surveys or administrative sources)
3. Feature generation
4. Algorithm or model choice
5. Combination of 3 and 4.

Combining difficult-to-explain machine learning models with deep features risks putting a black box on top of another black box; intuitive visualization tools can help mitigate this. The section on traditional PMT discussed the concern that PMT is a black box for policy makers and the public. The use of machine learning models that do not produce everyday variables with intuitive scoring (as in table 6.3) risks doubling down on this opacity. However, visualization tools have been developed that can help policy

Figure 6.6   Decomposition of Sources of Improvements in Model Accuracy: Models versus Data

Legend:
- Quantile linear regression on expert features (status quo)
- Gradient boosting on expert features
- Gradient boosting on expert and statistical features
- Quantile linear regression on expert, statistical, and deep features
- Gradient boosting on expert, statistical, and deep features AI based

X-axis: Inclusion error
Y-axis: Exclusion error
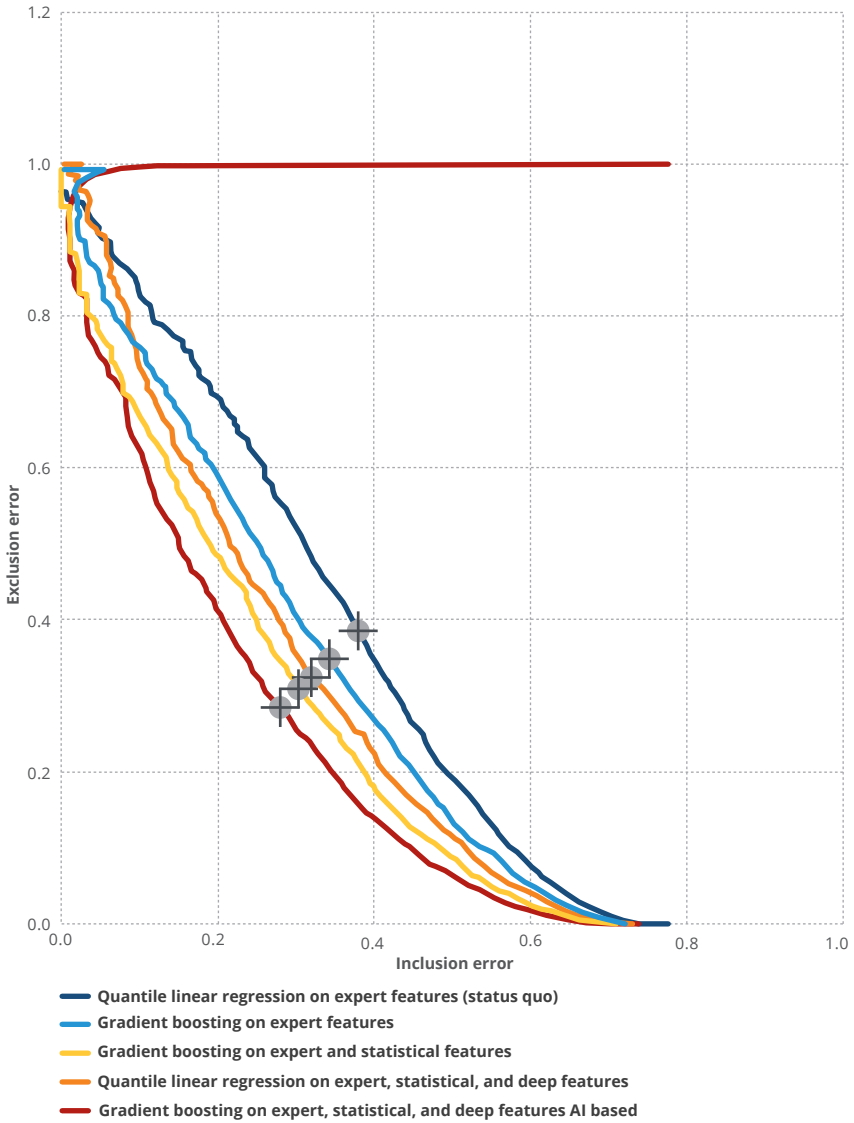
Source: Noriega-Campero et al. 2020, supplementary materials (https://dl.acm.org/action/down loadSupplement?doi=10.1145%2F3351095.3375784&file=p241-noriega-campero-supp.pdf.).

Note: AI = artificial intelligence.

makers and social workers understand why particular households are deemed eligible or not; figure 6.7 provides an example. The figure presents a situation in which a particular household has been scored in the broader income distribution and relative to the eligibility threshold and which of the household's characteristics are driving this, accounting for demographics, employment, access to services, and assets. This type of intuitive visualization tool can be useful for PMT in general and machine learning models in particular.

Nonetheless, machine learning–driven models and new variables are not a targeting panacea. Significant errors remain even when they are combined; when more labor income is directly measured, means testing and HMT offer more accurate eligibility determination. The Areias et al. (forthcoming) and Noriega-Campero et al. (2020) results show that while machine learning models may offer modest improvements when applied to standard data, it is the incorporation of new data into the models that is likely to drive improvements in accuracy. At the same time, significant errors remain. In Costa Rica, using machine learning and new data improved the inclusion and exclusion errors from around 30 to 24 percent; this is significant but not enough to mollify critics of PMT-based targeting. In Colombia, the inclusion and exclusion errors improved from around 38 to 28 percent, an even larger improvement but still leaving a significant fraction of the population misidentified. Improved PMT is still not a substitute for direct measurement of most or all income nor for interoperability and data integration.

Figure 6.7   Visualization Example of a Household's PMT Score and Its Drivers
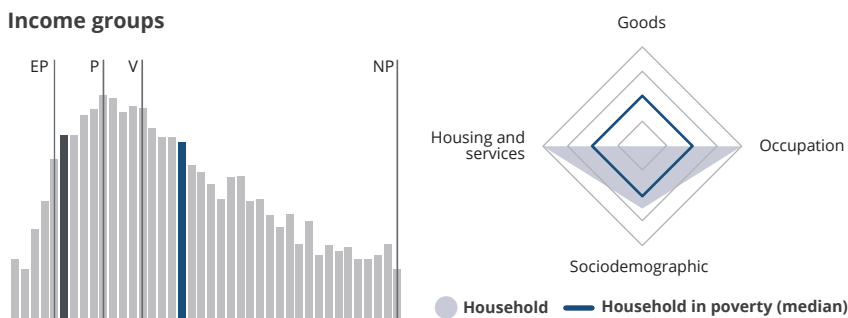
Source: Carrillo et al. 2021.
Note: EP = extreme poverty; NP = nonpoor; P = poverty; PMT = proxy means test; V = vulnerable.

Improving PMT with the New Big Data

The use of household- or individual-specific big data for improving targeting is generating excitement.[76] The prior section showed the modest results from applying new machine learning methods applied to traditional survey data. New proxies are needed—ideally, proxies that are related to household monetary welfare but not to the traditional proxies used in the models. Two main sources for these new data are public sector big data and private sector big data. The previous section showed how using the former—administrative data—as ancillary data can drive improvements in PMT; for example, in Costa Rica and Colombia. This section reviews how close social protection is to using private data for eligibility determination. The section also asks some bigger questions about whether some of these data should be used even if they are accessible.

A few recent and nascent experiences, Togo prominent among them, show how CDR data can be used for household-specific eligibility determination (beyond poverty mapping or proof of concept in research papers). In Togo, a collaboration between a team of academics, GiveDirectly, mobile phone operators, and the government used big data, including CDR data from private sector mobile operators, to target poor households in poor rural areas of the country (box 6.9). Similar work is being pursued in Bangladesh,[77] the Democratic Republic of Congo, Nigeria, and elsewhere.

BOX 6.9

Use of Private Big Data to Select Poor Areas and Poor Households in Togo

Togo is a small country in West Africa with high levels of poverty, even before the COVID-19 crisis. In response to the crisis and the economic pain of COVID-19-related lockdowns, the government immediately launched an emergency cash assistance program (Novissi), which provided electronic (contactless) transfers to nearly 600,000 informal workers in the areas most affected by the lockdowns and curfews, through the beneficiaries' cell phones. Individuals made their application to Novissi through a new cell phone interface created for Novissi. Eligibility was determined from a recently updated voter database, and transfers were made to informal sector workers (categorical targeting) in areas of cities where quarantines were most stringent (geographic targeting).

BOX 6.9 *(continued)*

Given the success of the fast response, a second phase was designed by the government in collaboration with GiveDirectly, expanding Novissi to rural areas, which, while not as affected by lockdowns, have the highest levels of poverty. To identify beneficiaries, an academic team[b] used big data to conduct a mixed geographic and call detail record (CDR)–based proxy means testing (PMT) approach (henceforth, the big data model). The team first used satellite and other geographic information system (GIS) data to develop poverty maps at the canton level, using the big data methods described earlier in this chapter in the section on geographic targeting. The estimated 100 poorest cantons were then selected to receive the Novissi transfers. The second step was to identify the poorest 10 percent of the residents in those cantons to receive the expanded assistance. To do this, the team conducted a phone survey of 8,900 individuals in these cantons to "ground truth" the models, matching the survey estimate of consumption of the surveyed individuals to their mobile phone data to enable the models to predict consumption from phone patterns, that is, allowing PMT modeling to use CDR as proxies to model consumption (CDR-based PMT). In the end, electronic transfers were made to 57,000 individuals.

A recently completed assessment compares the predictive power of the big data model in simulations, with several interesting findings:

- Compared with the sort of occupation-based (categorical) targeting of informal workers used in the first phase of Novissi and the straight geographical targeting being considered by the government as an alternative, the big data model worked better.
- The assessment also examined potential sources of exclusion and found that over half of the rural households surveyed were excluded from the program because they did not apply or were not able to register successfully using the cell phone application process, making these much larger potential sources of exclusion error than estimated by the model; 60 percent of those who successfully registered were deemed eligible by the model. Thus, only 19 percent of those surveyed were excluded at the model stage, highlighting the importance of outreach and ease of application.
- A pre-COVID-19 more face-to-face survey with more traditional consumption data could also be matched to phone records and allowed the simulation of a hypothetical nationwide program.

    ○ The big data model again outperformed straight geographic targeting and categorical targeting to the informal sector.

continued next page

BOX 6.9 *(continued)*

○ The big data model performed similarly to narrow, occupation-based (categorical) targeting to agricultural workers only, which a priori was known as the occupation with a lower income level among workers.

○ However, the inclusion and exclusion errors in the big data model were estimated to be 50 percent, compared with 37 percent for a traditional PMT using proxies available in traditional household surveys (as presented in the section in this chapter on traditional PMT models).

The assessment concludes that the big data model was the most accurate option available to the government at the time and for the emergency transfer, and that such models may be useful for humanitarian response when traditional data are missing or out of date, but in broader circumstances such models should be seen as a complement to traditional methods and not a substitute.

Sources: Aiken et al. (2021); Boko et al. (2020); https://medium.com/center-for-effective-global-action/using-mobile-phone-and-satellite-data-to-target-emergency-cash-transfers-f0651b2c1f3f; https://www.worldbank.org/en/results/2021/04/13/prioritizing-the-poorest-and-most-vulnerable-in-west-africa-togo-s-novissi-platform-for-social-protection-uses-machine-l. This box references unpublished and ongoing work, which is summarized at https://medium.com/center-for-effective-global-action/using-mobile-phone-and-satellite-data-to-target-emergency-cash-transfers-f0651b2c1f3f.

a. The academic team was based at the University of California, Berkeley, Innovations for Poverty Action, and Northwestern University. It is rare for eligibility determination for a government-led program to be done by unaffiliated institutions. It was done in this case with great care around data security and privacy, prompted by the crisis and more feasible because the financing was basically from GiveDIrectly rather than the government. The Agence Nationale d'Identification was recently created in Togo to collect information for the dynamic assessment of needs and conditions and eligibility determination for multiple social protection programs, under the supervision of the presidency. This will require both significant capacity building as well as establishing a regulatory framework that will (1) ensure that only the targeting institution could handle the data directly, and (2) ensure informed consent when people apply for the transfer program or in the way mobile phone operators enroll customers.

Interest in using big data for targeting stems in part from the prospects for reducing costs and improving the speed of response. In the case of Togo, such gains were certainly realized. Within a matter of weeks, contactless transfers were delivered to 57,000 individuals (the desired scale) in very poor areas of a poor country. This stands in marked contrast to the cost and timeline of mounting the traditional (partial) census sweeps or CBT exercises used for other programs. Moreover, recertification exercises would similarly require lower data collection costs (although they would still face issues of winners and losers). It seems likely that such gains should be widely replicable, although there is an important first step in gaining access to the data. In Togo, CDR data were obtained from the country's two telecom operators for a temporary program in the midst of a historic crisis. In a more normal setting, the issue of access might not be resolved so quickly, but at worst it should be a one-off, start-up project and the use of digital payment of benefits may provide an incentive.

The other reason for interest in using big data for targeting is the hope to improve accuracy by offering additional or better proxies, but it is wise to remember that they are still proxies rather than direct measurement of income. Big data will predict welfare with error, as panel a in figure 6.8 shows, plotting real welfare against that predicted by a CDR machine learning model for Rwanda (Blumenstock, Cadamuro, and On 2015). The nature of the approach of PMT with different proxies is shown when compared with panel b in figure 6.8, which plots real welfare against a

Figure 6.8   Scatterplots of Real and Predicted Welfare from PMT Based on CDR in Rwanda and Survey Data in Kenya
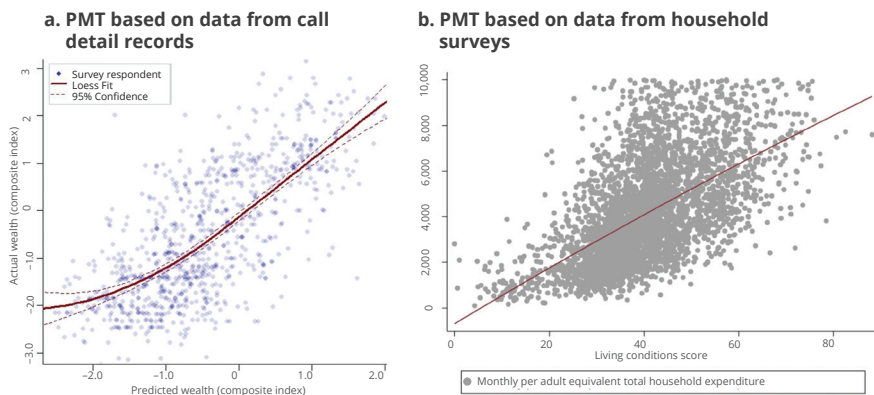
Source: Blumenstock, Cadamuro, and On 2015 for panel a; Silvia-Leander and Merttens 2016 for panel b.
Note: CDR = call detail records; PMT = proxy means testing.

PMT prediction based on traditional survey data in Kenya (Silva-Leander and Merttens 2016), a not very different looking picture compared with panel a. Aiken et al. (2021) use simulations to assess Togo's targeting, finding the CDR-based PMT to be more accurate than the other available options considered by the government, such as geographic or categorical targeting of informal workers. However, in a second simulation, they find that a nationwide CDR-based PMT is significantly less accurate than a PMT based on traditional data (which were not available for the actual targeting). The paper concludes that the new big data approaches can complement traditional targeting methods, particularly in crisis settings or where traditional data are incomplete or old. They call for more research to explore how big data can be best combined with traditional data to improve targeting. Given that CDR data might be related to household welfare in different ways than traditional data, there may indeed be potential for a combined PMT to be more accurate than either one alone.

To understand how accurate big data–based PMT can be requires a larger body of research and experience than has yet been generated. As the field develops in the near future, it will be important to keep an eye on how well it resolves some technical questions:

- First is *the unit of observation issue*, mapping between SIMs, phone numbers, individuals, and households, as discussed in the introduction to this chapter. As inexpensive phones become ever more ubiquitous (due to secular trends and in some places social policy initiatives to provide phones for the very purpose of enabling their use for social assistance and/or financial inclusion initiatives), issues of outreach and errors of exclusion should diminish, but issues of multiple SIM cards or phones per person or household may not. It may be possible for machine learning to combine multiple SIMs into a single individual based on common networks and patterns, or to identify common household members using nighttime locations, common networks, and calls between members, although this has not yet been seen. Thus, analysts and the programs they serve will need to find a way to get a comprehensive composite of the individual or household from the data, or acknowledge that they may face errors of inclusion as the phone use of people with multiple phones will be underestimated. The extent of such errors and where in the income distribution they become large enough to be of policy concern are empirical questions subject to investigation.
- Second, *there is a question of what "ground truth" or training data to use to build the model.* In the ideal, ground truth data should provide an accurate

measure of welfare that reflects current reality. There are various challenges to that:

o Big data–based PMTs are especially attractive where traditional data are scarce, but this poses an issue of training or ground truth data. As noted at the start of the chapter, many of the big data models are trained with Demographic and Health Surveys. Although it is widely available, the Demographic and Health Survey wealth index is itself just a proxy for the underlying welfare measures for which the analyst must find proxies.[78]

o If it is desired to develop big data–based PMTs for use during a crisis, there are two choices, both of which have drawbacks, and there is not yet a body of work to demonstrate their size or suggest which approach is more robust.

o If models are developed based on matching big data with consumption or income collected in normal times (for example, from the last comprehensive household survey), the models may not be robust to crisis-induced changes in behavior. During the COVID-19 stay-at-home orders, for example, people moved around less, and, generally, people who could work from home and afford to stock up on food stayed home more than those whose jobs demanded that they go out to work and shop daily.[79] People may also have used their phones more because they wanted to gather information about people/settings they could no longer visit, or to conduct more transactions by phone; or they may have used them less because they had lost income or jobs and could afford less usage or needed them less for their work. The sizes of these biases were initially uncertain and there were hypotheses in countervailing directions.

o If models are developed using welfare data collected during the crisis, the patterns of use may be correctly mapped, but the crisis may impede the collection of the welfare data, and noisy welfare data will impair the precision of the model. Gathering data during a crisis is increasingly being done by phone rather than face to face, to allow for speed and, in the case of a pandemic or conflict, to lower risk, and in the case of a natural disaster, to reduce issues around interruptions to transportation networks. However, phone surveys typically have much shorter interview times than full Living Standards Measurement Study surveys or Household Income and Expenditure Surveys, with truncated questions on income or consumption or alternate definitions of welfare, and can have issues of which respondent in the household is interviewed, all of which make sound measurement of welfare problematic.[80]

Using big data for eligibility determination may raise social or ethical concerns that bear consideration. The use of data on personal actions, such as phone calls or social media posts, for eligibility determination takes on a larger cultural dimension of "Big Brother watching," especially if people do not fully understand that it is not the content of the conversations that is monitored but the somewhat less intimate details of their frequency, length, origin, whether incoming or outgoing, and text or voice (although it does include who else you have been talking to). Will people feel comfortable if how they chat with friends by phone, Facebook, or Twitter; search on Google; or buy on Amazon influences their eligibility for social protection programs? And how will the outcomes be explained? Although this book is not the place, we believe it is important that a deeper look be made into not just what legal protections are required governing data ownership and protection, but more broadly what sociocultural considerations should be explicitly brought to bear on these issues.

Big data and machine learning may raise some significant human rights issues (see box 6.10), although work is underway to try to provide privacy guarantees while still allowing public good applications.[81] In the case of Togo, the academic team carefully tested for demographic parity and

---

BOX 6.10

Machine Learning, Big Data, and Human Rights

Machine learning, private big data, and biometric technology may generate gains to the delivery system and selection of beneficiaries of social protection programs, but they also pose risks and challenges that must be considered and minimized or mitigated.

Among the advantages, biometric identification systems help in uniquely identifying people for social protection through systems such as fingerprints, iris, and face recognition, allowing not only such identification but deduplication and interoperability of systems. Machine learning and big data can help in understanding poverty and patterns of need, matching people to programs, and changing the interaction between people and the state (Gelb and Metz 2018).

However, there are multiple risks that have implications for human rights. Ohlenburg (2020a) and Sepulveda (2018) are thoughtful sources. Among the risks they cite are (1) inaccuracy of data or exclusions from data or algorithms, (2) identity theft/data protection, and (3) security risks and the misuse of data.

BOX 6.10 *(continued)*

With biometric technology, data inaccuracy may occur when individuals enroll their biometric data. For example, biometrics are not good at reading the fingerprints of a share of manual workers or the elderly as finger pads can be worn down, or the irises of those with glaucoma or cataracts. Thus, during the process of matching an individual's biometric against a template stored in a database, inaccurate matches/nonmatches may occur.

Machine learning algorithms also risk perpetuating inequalities and bias (exclusion) against certain groups.[a] Modeling may be based on data that reflect historical biases. When a learning algorithm is presented with data that reflects historical discrimination, it may learn to imitate the biased patterns of the past. For example, a machine learning algorithm could discern that variables such as ethnicity are good predictors of outcomes. But except for certain affirmative action programs, in most places, ethnicity-based targeting would not be acceptable. Or the machine learning algorithm could pick up other aspects of historical discrimination, for example, related to ethnic ghettos or redlining a discriminatory practice in which services are withheld from potential customers who reside in neighborhoods classified as "hazardous" to investment; these residents largely belong to racial and ethnic minorities (Zou and Schiebinger 2018). As computers cannot distinguish between ethical and unethical decisions, data scientists need to be aware of historical biases and consider how they can be addressed.[b] Machine learning will exclude all that are not part of the data-generating technology on which the artificial intelligence system relies; mobile phone use is an obvious example. As access to digital services tends to rise with income, the poorest are the most likely to be data poor as well and consequently excluded.

A related risk is that machine learning algorithms learn best to predict the data set on which they are trained. The *World Development Report 2021* on data highlights several issues (World Bank 2021b). For example, an algorithm to predict drug use in California trained on police arrests and predicted predominantly African American communities despite survey data indicating widespread use across Caucasian and African American communities (Smith IV 2016). Facial recognition algorithms were first trained on Caucasian males and so recognize them best; they perform less well on African Americans (Hill 2020) and worst on African American females, with error rates reaching 35 percent (Buolamwini and Gebru 2018). In the case of the former, this has led to mistaken identities and arrests.[c] Voice recognition also suffers from male bias (Tatman 2017).

BOX 6.10 *(continued)*

For data protection, it is necessary to have proper safeguards in place to ensure that identifier indicators are hard to compromise, reducing the chance of imposters gaining access to data. The purpose of social registries is widespread use by different government agencies, and a strength is that they can draw together data from different agencies. However, this richness carries risk. Without proper safeguards, systems may allow too many users too much access. Data protection and security must be core elements of system design, starting with the internal procedures of organizations storing information and with proper controls.

Moreover, as *World Development Report 2021* notes, in principle, data protection laws limit the use of personal data, but generally exceptions exist. In most cases, these are limited to specific uses, such as national security, but in other cases, they are wide ranging. Justifications for these exceptions are required in a third of high-income countries but less than a tenth of low-income countries—which opens the door for additional opportunities for unchecked state surveillance, thus undermining trust in data use (Ben-Avie and Tiwari 2019). In addition, *World Development Report 2021* notes that the increasingly widespread practice of linking data sets stretches the limits of anonymization, creating the possibility of reidentifying deidentified data and blurring the boundary between personal and nonpersonal data (Lubarsky 2017).

On security risks and misuse of data, manipulation of personal data raises the risks of violations of rights, such as: (1) loss or unauthorized access, destruction, modification, or disclosure of data; (2) misuse of the information by governments or the private sector for systemic surveillance of individuals; and (3) vulnerability to hackers. The use of facial recognition in the context of the governments' ability to curtail rights such as freedom of assembly and expression through the identification of protesters is a particularly worrisome concern.

a. Lindert et al. (2020), Sepulveda (2018), and Sepulveda and Nyst (2012) highlight that the exclusion of people from social programs or from obtaining IDs is, among others, due to lack of awareness of the enrollment; limited infrastructure or presence of the enrollment office or station, mainly in rural and the poorest areas; physical mobility of individuals; cost or any other administrative requirement; physical inability to provide reliable biometric information; and cultural barriers and gender norms.

b. Ohlenburg (2020a) indicates that when a learning algorithm is presented with data that reflect historical discrimination, it will learn to imitate the biased patterns of the past, hence perpetuating the historical discrimination.

c. See, for example, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html and https://www.cnn.com/2020/06/24/tech/aclu-mistaken-facial-recognition/index.html.

systematic prediction errors for a large number of different subgroups, such as gender, ethnicity, and religion, but without careful management, oversights, and audits, machine learning algorithms can encode bias.

It will also be important to learn whether or how much using big data to determine eligibility for benefits may change behavior. If financial transactions are used to model household welfare, will some transactions move back to cash to hide them? Will households avoid purchases of important goods just to remain eligible? Will households change or mask their calling behavior or even their movement patterns? Will they self-censor on social media or post misleading information? For example, when it was realized in Kenya that GiveDirectly was determining household eligibility based on satellite assessments of roofing, households would not upgrade their roofs. As credit based on mobile phone use systems has scaled, manipulation has become commonplace as borrowers learn what behaviors will increase their credit limits (see the references in Björkegren, Blumenstock, and Knight [2020]). Potential incentive issues are not unique to this type of big data,[82] but they take on new forms that are not yet well explored or quantified. Efforts are being made to make machine learning approaches more robust to strategic behavior (Björkegren, Blumenstock, and Knight 2020; Hardt et al. 2015).

In summary, private big data offer exciting possibilities for improving eligibility determination, but work on various fronts—technical, social, and legal—will be needed for fully dimensioning and grasping these. This is a field that has been developing quickly with big advances even during the time the rest of this book was being conceived and written. It seems likely that data science will advance as quickly as ground truth data can support, and this will induce further needed attention on the regulatory and policy fronts.

Key Elements for Community-Based Targeting

Traditional CBT takes a far different tack on discerning who is poorer. It forgoes interoperability among government databases, household-by-household quantitative surveys, and fancy algorithms. Rather it uses a group of community members or leaders, en masse or in committees, as the main agents in the selection of beneficiaries for social assistance programs. The community members are expected to have enough knowledge about their neighbors from their day-to-day lives—who buys how much of what in the market, how people work, what clothes or shoes they wear, and how they participate in community social interchange—so that they could do some sort of needs assessment without carrying out special purpose data collection. In the sense that the data used are already generated for other purposes (community life), it is like big data but maybe the term

traditional data strikes the right chord. Common techniques used in community targeting exercises are community or participatory wealth ranking[83] (Kebede 2009; Zeller, Feulefack, and Neef 2006), participatory rural assessment (Chambers 1994), and HEA (Holzmann et al. 2008). Traditional CBT methods not only rely on local information, but may incorporate local notions of deprivation into the selection criteria of social programs.

There is an extensive literature on CBT, for example, Conning and Kevane (2002), Himmelstine and McCord's (2012) annotated bibliography of more than 100 studies, and McCord's (2013) distillation of many CBT experiences. In this literature, there is significant accord on the potential importance of the information base of communities and the process of involving them. Handa et al. (2012), for example, show that CBT in East Africa had on average better results than Coady, Grosh, and Hoddinott (2004) found in their benchmarking across many methods and programs. The literature also shows that when it is effectively implemented, CBT can generate widespread program support even if only a portion of the population benefits. There are also frequent citations of a set of challenges to be managed with respect to community dynamics—how to reduce risks of errors of inclusion stemming from elite capture and general data manipulation, how to minimize errors of exclusion from any systemic patterns of exclusion in community life, how to minimize any friction caused by drawing distinctions, and asking community members to help in that process. Alatas et al. (2019) distinguish between formal and informal elites. While they find evidence of formal elite capture (although relatively modest), they find no evidence of informal elite capture. Another theme of the literature is that while in general the method produces progressive outcomes, there is a lot of variation in both outcomes and how CBT methods have been implemented from place to place. For example, Premand and Schnitzer (2018) show that in Niger, CBT was done with three committees, and the results were then triangulated. The authors show that doing CBT with just one committee would lead to manipulation and suffer from information asymmetries. However, using three committees and triangulating across them reduced these risks (although not fully). Hence, on variations of outcomes and implementation, CBT is like other targeting methods.

Without trying to repeat the classic references, this section mentions some of the practical issues that have arisen in recent implementations of social assistance where communities have an important role, although sometimes not an exclusive one, in determining eligibility. These are illustrated with short details from various programs and a more detailed story on the use of CBT for Ethiopia's rural Productive Safety Net Program in box 6.11.

BOX 6.11

Ethiopia's Rural Productive Safety Net Program: Community-Based Targeting

The Ethiopia Rural Productive Safety Net Program (PSNP) uses community-based targeting to choose the households within the districts (*woredas*) and subdistricts (*kebeles*) where the program works up to the caseload allocated to the area. (Box 5.3, in chapter 5, explains geographic targeting.) The process has undergone continual scrutiny and the program has worked on improvements, highlighting a core message of this book—that targeting is not perfect or easy, but effort can often improve it.

In the PSNP, the process of selecting beneficiary households is run at the community level, carried out by the Community Food Security Task Forces, which are specified in the project implementation manuals as being comprised of a development agent, a health extension worker or volunteer community health worker, a Community Care Coalition representative, two or three elected female representatives, two or three elected male representatives, an elected youth representative, and an elected representative of the elders. The job of these task forces is to (1) identify eligible participants based on guidelines and training received from the kebele (see table B6.11.1); (2) identify households that can participate in public works and those without labor or other support that will need direct support; (3) display the proposed list of participants in public for at least a week, for comments and endorsement by the general meeting of the village residents; and (4) finalize the list and pass it to the Kebele Food Security Task Forces for verification and further action.

The early years of implementation experienced some common teething pains, which were addressed through capacity building as the program developed. The caseload identified was initially higher than the rationed allocations, but the program accommodated this with an increase from the originally planned 5 million to 8 million beneficiaries. Understanding and implementation of the guidelines were initially not uniform, but they improved in subsequent years after increased communication with regional and district staff and communities. In the early years, there was some tendency to dilute benefits through incomplete listing of all family members (so that more families could fit within the rationed caseload) or families being rotated in and out of the program within a year. Both of these were reduced when the automated Payment and Attendance Sheet System was introduced. Committees to

continued next page

BOX 6.11 *(continued)*

Table B6.11.1   Criteria for Selecting Beneficiary Households

Basic Criteria for the Selection of Households

The following basic criteria should be used to select households for participation in the safety net program:

- Households that are members of the community
- Chronically food insecure households that have faced continuous food shortages (usually three months of food gap or more) in the past three years and that have received food assistance prior to the commencement of the PSNP
- Households that suddenly become more food insecure as a result of a severe loss of assets and are unable to support themselves (over the past one to two years)
- Any household without family support or other means of social protection and support.

Criteria for Refining the Selection of Households

Having made the initial selection based on the basic criteria, the following factors should be examined to verify and refine the selection of eligible households:

- Status of household assets: land holding, quality of land, food stock, and so forth
- Income from nonagricultural activities and alternative employment
- Support/remittances from relatives or the community.

Sources: MoARD (2006); Van Domelen and Coll-Black (2012).

handle grievances and appeals were established in the third year of the program, with additional efforts to strengthen transparency within the process in the later years through, for example, posting beneficiary lists and introducing social accountability mechanisms through non-governmental organizations.

By the fourth year of program implementation, the evaluation showed that the program was well targeted within woredas (see table B6.11.2) and 85 percent of the survey respondents deemed the process to be fair. Periodic impact evaluation reports from 2006–14 show that PSNP benefit rosters have been to some extent dynamic, with both entries and exits, and 70 to 80 percent of the households from the first year of each evaluation period were still enrolled in the third year. The overall targeting of the PSNP is progressive and in line with results from other countries' public works programs. Moreover, it was the selection of poor households within participating woredas that drove those results much more than the geographic targeting, as illustrated in box 5.3 in chapter 5.

In 2020, the government of Ethiopia decided to revise the community-based targeting process going forward. The program will

BOX 6.11 *(continued)*

Table B6.11.2   **Household Consumption, Assets, and Land Access in PSNP Woredas, by PSNP Beneficiary Status, 2008**

Economic characteristics	Direct support	Public works	Non-PSNP
Total consumption (birr) per month, average	627	1,012	1,111
Land (hectares), average	1.0	1.1	1.4
Assets (birr), average	2,349	4,568	6,480

Source: IFPRI/CSA 2008 Household Survey as quoted in Van Domelen and Coll-Black 2012.

Note: Consumption refers to the value of total consumption (food and nonfood) both purchased and the value of production of self-produced goods. Assets are the value of livestock and productive equipment used in agriculture. PSNP = Productive Safety Net Program. US$1 = birr$9.19.

establish standard rules for a biennial assessment (or recertification) of all beneficiaries who have been in the program for more than three years and an assessment of households that benefited from livelihood services two years after they obtained funding for their business plans. Every four years there will be a major household retargeting exercise to correspond with the reallocation of the caseload. Moreover, the program is upgrading its heretofore rather modest management information system to be the seed of a richer household registry so that multiple programs in Ethiopia can share information.

Recent years have seen an increase in conflict and civil unrest, leading to large-scale cases of internal displacement, including of PSNP beneficiary households, necessitating the introduction of protocols on how to safeguard the entitlements of affected beneficiaries. In PSNP woredas affected by unrest, beneficiaries who become internally displaced will remain on the payroll for at least one year after displacement. Exiting or replacing displaced beneficiaries will not be permitted during that period. Displaced beneficiaries who return to the PSNP woreda within one year will be entitled to receive all the benefits from the absence period. Temporarily displaced people who are hosted in PSNP woredas may be assisted through the woreda contingency budget based on the community's assessment of their eligibility. In situations of protracted displacement within PSNP host communities, for at least two years, woredas will be allowed to relax the three-year community membership rule to include displaced families in the program if they meet the poverty/vulnerability eligibility criteria.

Sources: MoARD 2006; Van Domelen and Coll-Black 2012; World Bank 2020.

Differences in Concepts of Poverty

There is a tension in some social assistance programming with central planners often thinking in terms of money-metric measures of poverty whereas individuals and communities have their own interpretations of need and of who should benefit from social programs. This is a feature that runs through the literature. In the studies on Indonesia and Niger cited in chapter 5, communities seemed to rank on potential to earn income, not just on actual consumption. In Rwanda, Nizeyimana, Lee, and Sim (2018) use an experiment to compare the Ubudehe classification system, which is a community wealth index that groups households into different socioeconomic categories,[84] with a sort of wealth index[85] as well as money-metric poverty and consumption. The authors find that the predictability of Ubudehe categories using the set of variables, household assets, and consumption was only 35.9 percent. In Zimbabwe, Robertson et al. (2013) analyze a pilot cash transfer program for vulnerable children and find poor agreement between the Participatory Wealth Ranking used to target the program and a wealth and asset-based index constructed from a census of households used for the evaluation. In contrast, in Bangladesh, Feulefack and Zeller (2005) compare results from a Participatory Wealth Ranking with data from a Living Standards Measurement Study survey–type questionnaire for measuring monetary poverty. They show that Participatory Wealth Ranking scores compared with money-metric measures are in significant accord: up to 8 in 10 households are correctly predicted as to whether they live in extreme poverty or not.

Sometimes the center tries to steer the concepts of poverty to be used, leaving less autonomy for the community in that respect but still bringing to bear their local knowledge. In this case, training and capacity-building activities are core for harmonizing concepts across communities and to guide communities toward the main goals of the project. But this choice also depends on program objectives and what the program aims to achieve. In Mali, after running the first wave of identification of 5,000 households, monitoring revealed that small families with children were being included in surprisingly small numbers. The community committee members were retrained to have a better understanding of the process and intended concepts for selection. In the Republic of Congo, the National Statistics Office, with all its professional experience, provided training to the local social workers and community members so that they could understand the criteria the government wished to use to determine which households should be prelisted for further interviews for the social registry. The traditional criteria in use by the community tended to favor demographic vulnerability, whereas the program was more based on money-metric criteria. In Malawi, Basurto, Dupas, and Robinson (2020) show that the chiefs did not

select individuals based on poverty; they selected households that ended up having larger returns from the program.

Challenges to Effective Rankings

One challenge faced by traditional CBT is related to the ability of the community members to rank individuals from the worst to better-off (irrespective of the exact concepts used) due to the quality of networks and information asymmetries. In Western Honduras, Bergeron, Morris, and Medina Banegas (1998) experimented with running participatory wealth ranking workshop sessions in different communities, with members of organized small farmers' groups, to rank the potential beneficiaries of an agricultural development project according to food security. The participants, who generally knew each other well, were split into small sets and each set was asked to rate the food security status of all the households in their organized group. The findings show poor agreement among the raters, and the authors suggest that people have different knowledge and understanding of (1) what is the subject they are rating, and (2) the criteria. Moreover, a few people ended up dominating the discussion, leading to bias. In Premand and Schnitzer's (2018) work in Niger, individual raters were found to know about three-quarters of the households they were asked to rate in their village. The results show that there would be substantial exclusion errors due to imperfect local knowledge in rankings from a single local committee, but triangulation of the results across three committees substantially addressed the problem, which was treated by proper implementation.

Even when there is sufficient knowledge, effort or fatigue may be an issue. In Indonesia, Alatas et al. (2012) show that the meetings to rank households ran longer than one and a half hours and communities had to rank 54 households on average. The authors show that for households presented earlier in the meeting, the villagers' rankings were relatively accurate at identifying the poor, but the accuracy worsened as the meeting progressed, suggesting that fatigue might have set in and undermined the value of including the community in eligibility decisions. An analogy to the span of attention in committee meetings is presented by Banerjee et al. (2010). The authors present evidence that people are preoccupied with more immediate needs and may not have the attention span needed to monitor providers coming from village meetings in India in which parents were encouraged to question village education committees and local government officials about education.

The structure, inequality, and size of a community are relevant for pure CBT performance. Population size, geographic dispersion, inequality among residents, and distance between villages matter for targeting outcomes.

Communities with high levels of inequality generally produce better outcomes. Larger communities have less cohesion and consequently more limited knowledge of members' status and needs, which leads to worse outcomes. These problems may be exacerbated in communities that are less well established or more fluid, with transient rather than well-established populations. For example, in the Republic of Congo's Lisungi program, assessment of the first phase of the program indicated worse performance of community committees in urban areas. Where communities are less homogeneous in social, cultural, ethnic, or racial terms, CBT may be less effective, particularly in contexts where access to resources is contested and there is instability or latent conflict between different population or ethnic groups. Nevertheless, community participation helps legitimize the process in many contexts, as presented in Alatas et al. (2012) and Premand and Schnitizer (2018).

An interesting finding in the Democratic Republic of Congo can help in understanding the importance of community participation in the process, even when a village chief is in charge of ranking/listing the beneficiaries. An impact evaluation conducted for the Eastern Recovery Project[86] measured the accuracy of beneficiary selection done by village chiefs. The initial findings from the baseline of the randomized trial were that the chief's selection could be as good as the choices made by community committees. For this exercise, the findings indicate that using the chief rather than communities was more cost-effective, mainly because it took less time than using the traditional committees. Nevertheless, the chief's selection was endorsed during a large community meeting, and more research is being conducted to measure the political repercussions as well as elite capture/nepotism that might mar the process. This approach of having the village chief's list validated by the community may be an effective alternative to training and using community committees. However, it is important to highlight that in some places, the chiefs refused to select beneficiaries for fear of social discord and voluntarily reverted back to community targeting or public lotteries.

Setting Up the Community Committees

The composition of the community committees can rely on existing local institutional structures such as semiformal or informal village councils, school boards, mosques or parish councils, organized nongovernmental organizations, or traditional leaders, assisted perhaps by institutional organizers. These may be supplemented by ad hoc groups or full community mobilization exercises.

How community committees are set up can help ensure that some of the challenges of CBT can be managed. For example, if there is concern that

any committee members will manipulate the process to capture benefits for themselves or their dependents, having multiple committees cross-check each other's work can be helpful. If there is concern that certain groups may be excluded due to cultural norms or power structures, the social assistance program may establish rules for ensuring the inclusion of such groups by adding explicit rules and quotas for their representation in the community committees charged with selection of beneficiaries and other roles, and/or it may establish minimum quotas for the groups in the program.

The community committees must fit into the hierarchy of rules and responsibilities of each organizational/geographical level in a clear way. There are different levels of committees, some as committees of officials and public servants that have only a supervision/monitoring role and others such as the committee of local leaders that is actually in charge of selection, depending on the country's administrative/geographic partition, but their roles and responsibilities generally fit a pattern.

The main committee for CBT is at the lowest level—the community, municipality, and ward-level committees. These are chosen by the community to work as the main agents to define the eligibility of beneficiaries. They are composed of representatives of community leaders and other members from traditional groups, including women, youth, school boards, and mosques; civil society members; village councils; and traditional leaders. Among other things, they are responsible for communication; helping in preidentifying poor households; preparing and validating lists of beneficiary households at the level of social action sectors; participating in educational and information campaigns; being a point of contact for the community to authorities for information on the processes of identification, payment of households, and compliance with conditionalities; and collecting complaints. Their main role is to strengthen communication about the program at the grassroots level, providing a space for dialogue between beneficiaries and the high-level administration. When communities are large, a subcommittee at the lowest level, such as the village or neighborhood level, is created. The main function of this lowest level committee is to help the main community committee by calling meetings and spreading program information. This committee may also help and validate the geographic premapping of all the households in the village or prelist the subset it deems neediest when the community committee is in session for validation of listing. Hence, the lowest level committee is important as a liaison between the community and the villages, ensuring that some village households are not excluded from the process due to poor information, bias, or discrimination.

Other committees can be established to provide some political support for CBT. They are composed mainly of officials, and they are in place for monitoring and supervision of programs. Even for other targeting methods, these committees can be created. The following two levels are often seen:

(1) *Decentralized level of the official government.* An intermediate-level committee may monitor the activities of the community and village committees. It is composed of representatives of civil organizations and officials from the ward/county level to support the monitoring of the program by supervising community-level committees.

(2) *Departmental, district, or state-level committee.* This type of committee would oversee and support project implementation, review and approve project implementation reports, rule on cases of litigation involving local monitoring committees, supervise other local-level committees, and so forth. Often such committee is composed of government authorities from the social sectors involved, such as education, health, agriculture, sanitation, gender, civil society, and elected officials. It includes dedicated project officers plus other members of the government, such as the head of the social workers, head of the district school board, and so forth.

The following are some examples.

In Mali, the Jigisemejiri program has the following three-level structure:

(1) Community, municipality, and ward-level committees consist of representatives of relevant social sectors (health, agriculture, social services, and education) at the local level, local elected leaders, civil society representatives (for example, women, youth, and the elderly, and local health agents), and nongovernmental organizations.

(2) Decentralized-level of official government, which consists of officials at the subprefecture level, or their representatives; decentralized-level officials from the Ministère de la Solidarité et de Lutte Contre la Pauvreté (Ministry of Solidarity and Humanitarian Action); an agent recruited by the program; officials from decentralized services of the Commissariat à la Sécurité Alimentaire (Food Security Commission); officials from different national directorates that are concerned with the program, such as Health, Social Protection and Solidarity, Basic Education, and Women, among others; as well as members of national civil society organizations, such as the National Federation of Community Health Agents, the National Center for Promotion of the Economy and Solidarity, and a few other centers.

(3) Departmental, district, and state-level committee members consist of officials such as the mayor or the mayor's representative, directors of the Regional Directorate of the Ministère de la Solidarité et de Lutte Contre la Pauvreté (Ministry of Solidarity and Humanitarian Action), and members of same institutions described in part 2.

In the Republic of Congo, the Lisungi Program has the following three-level structure:

(1) Community, municipality, and ward-level committees consist of the head of the social sector in the Cisconscription d'Action Sociale (Local Social Assistance Offices) (CAS); CAS social workers; local representatives of women's, youth, and elderly groups; and representatives of civil society in areas such as the National Network of Indigenous People (to represent this population in the Likouala region), who helped identify potential beneficiaries of the program.

(2) Decentralized-level official government representatives consist of officials at the subprefecture level, the head of the CAS, school and health district directors, the head of department of the police force, the heads of health committees, other key actors in relevant sectors, at least one representative of civil society working in the district, and one member of the National Network of Indigenous Peoples of the Congo in the Likouala region.

(3) Departmental, district, and state-level committee representatives consist of department-level officials, such as the mayor or the mayor's representative; department directors in the Ministère des Affairs Sociales et de l'Action Humanitaire (Ministry of Social Affairs, Humanitarian Action, and Solidarity) and for all other relevant sectors such as health, education, and agriculture, among others; at least two representatives of civil society working in the department; and one member of the National Network of Indigenous Peoples of the Congo in the Likouala region.

Evolving Roles and Mixing Methods: CBT and PMT

As government-led social protection systems have been developing and building capacity over the past decade or two, the ways that communities have been brought into targeting processes have been evolving. Indeed, a certain fuzziness is developing about what is meant by community targeting as the roles assigned and other information and processes brought to bear vary a lot. Often community members or committees are heavily involved in processes that also involve a PMT. Community members may help, through a structured participatory exercise, to organize a prelist of households for whom information for a registry or a PMT is gathered. The size of the prelist is determined by the resources available and can rely on geographic information to define local-level quotas using a geographical targeting approach. Sometimes the community in large meetings or members of a community committee provide validation/challenge to the list of potential beneficiaries from the PMT. After the determination of scores,

communities can also help in identifying a predetermined number of mis-classifications to be reassessed.

The following examples illustrate the variations.

In Senegal, the Programme National de Bourses de Sécurité Familiale (Family Security Grants Program) aims to select households in extreme poverty using a three-stage approach. First, geographic targeting is used to determine quotas at the lowest geographical level. Second, a CBT identifies potential beneficiaries of the social program (listing) to be surveyed. Third, a PMT is administered. The CBT committee in charge of this prelist is orga-nized based on existing structures (the elected village representatives or the elected neighborhood committees) and local religious and civil society representatives.

In the Republic of Yemen, the Social Fund for Development uses the existing networks in the community for outreach and to deliver programs during conflict. Prior to the current conflict, the Social Fund for Development had created effective administrative and delivery systems, relying on community-based organizations and local councils, tribal structures, religious-based groups, civil society organizations, private sector organiza-tions, informal networks, local councils, or other formal local institutions. The community-based network helped to select and implement public works projects and identify beneficiaries for both labor-intensive public works and direct cash transfer programs, working as the method for house-hold selection. The involvement of communities in the Republic of Yemen paid off according to Al-Iryani, de Janvry, and Sadoulet (2015) since evalu-ations indicate that the majority of the communities and households that were interviewed considered that the projects met their priority needs effectively.

In Kenya, the Inua Jamii cash transfer program is managed by the Ministry of Labour and Social Protection and the Ministry of Devolution and Arid and Semi-arid Lands and uses the community in the initial and final stages of beneficiary selection. Locally, the program involves the staff of local county and subcounty offices and a network of Beneficiary Welfare Committees in each location. The central level sets the rules, but the county coordinator provides administrative support to the subcounty offices, which support the implementation of cash transfer programs and engage-ment with beneficiaries. Chiefs and assistant chiefs act as Inua Jamii ambas-sadors and information resource persons for beneficiaries and their caregivers by regularly organizing public, open-air meetings (*barazas*) to provide information on the cash transfer programs. Beneficiaries are identi-fied by subcounty offices in a participatory process with the communities that includes community sensitization, engagement activities by chiefs and assistant chiefs, and barazas to inform potential beneficiaries about the

eligibility criteria of a cash transfer program. At the end of this process, a prelist of potential beneficiaries is ready to be screened through the PMT. After registration and application of the PMT by the subcounty officers, barazas are again held to validate and confirm the final list of eligible beneficiaries. Beneficiary Welfare Committee members are also agents for grievance and redress as they are the first contact points to respond to beneficiaries if they seek any information, have complaints, or need to update their information/record with the cash transfer program. Moreover, the Beneficiary Welfare Committees organize meetings and information-sharing sessions to educate cash transfer beneficiaries about their rights, responsibilities, and entitlements.

In Malawi, the Unified Beneficiary Registry, which is used to determine eligibility for the main cash transfer program and may serve others in the future as well, has built its implementation process around district and community structures that all exist and operate organically and independently of the Unified Beneficiary Registry. District-level actors cover the roles of coordination, training, and supervision, while field implementation is carried out by Area Executive Committee members,[87] Community Social Support Committee members, and community leaders. The Community Social Support Committees are composed of community members chosen by the community to work on programs, and group village heads are in place to support the implementation of the Unified Beneficiary Registry. The Area Executive Committee conducts a community meeting to elect the Community Social Support Committee, which then validates the prelisting of 50 percent of the households to be interviewed at the village level. Home visits and interviews are carried out by the Area Executive Committee members, with a Community Social Support Committee member serving as the liaison to the community and assisting the Area Executive Committee team in locating the households on the list. After data collection, the Area Executive Committee conducts the second community meeting, in which the Community Social Support Committee, other respected members of the community, council members, group village heads, village heads, and community members (including those interviewed) are all invited, to validate the PMT ranking and identify any households they believe were misclassified, excluded, or have appeals cases. Each household is discussed, and the list is updated with a community validation column that notes whether the community is in agreement (and its reason). Households that want to appeal their status or believe they were erroneously excluded from the list can be interviewed and registered by the Area Executive Committee member. Finally, a second-round PMT is applied to the updated Unified Beneficiary Registry data that contain the updated information on the new households and the Unified Beneficiary Registry list is ready for use by social programs.

In the Philippines, although the Listahanan social registry is usually classified as based on a PMT with a periodic survey sweep, the community also has a role. After the door-to-door survey sweep, application of the PMT, and compilation of data, the program administration posts the preliminary list of the poor in the community. Some community sessions are held to get feedback on the validity of the list. Any household that the community identified as missed by the registration wave could then be surveyed. Discussions also cover how well the poor/nonpoor classification jibed with local perceptions and identify inclusions or exclusions from the list. All feedback is received and recorded in the Complaints Form and a Local Verification Committee—which currently consists of the Local Chief Executive or his/her duly authorized representative, the Local Social Welfare and Development Officer, the Local Planning and Development Coordinator, and at least two representatives from nongovernmental organizations—is in charge of resolving all grievances and appeals received.

Support, Information, and Capacity Building

Although CBT is sometimes appealing because it takes advantage of existing local knowledge and informal leadership structures, good CBT requires information and capacity-building activities and physical investment. Proper information and training sessions are necessary to support communication with the committees about the program, its rules, and the layers of responsibilities of all actors. In many cases, some facilitators specialized in CBT procedures go from village to village to support the process. Training activity, such as how to inform people about the program, how to deal with grievances and complaints, how the payment mechanism works, and so forth, is core for bringing knowledge about the program to community members. Information sessions covering program objectives and procedures to identify beneficiaries are needed in all variants. When local communities are involved in collecting household information, training needs are even greater. At the same time, communities must be equipped to run the processes they are assigned. Stationary, information technology equipment, transportation, remuneration, and so forth should be in place to enable community committee members to carry out the roles assigned to them.

Although it is often unremunerated, community labor is not free. Program planners and administrators need to consider the costs,[88] even if uncompensated, of the time of the community members or leaders in the various roles they may play in helping to carry out program functions. There is likely an opportunity cost for the time in the traditional economic sense. There is also the potential for the role with respect to the program to affect various roles the person already holds, in ways reinforcing or undermining.

For example, a teacher may be involved in selecting poor children for scholarships, but also must interact with the parents of all students on their learning, including those of children not selected.

Clear protocols for oversight and control procedures are needed even when communities are given significant authority in beneficiary selection. A regular monitoring and evaluation system, which includes spot checks, process evaluations, and independent audits, will need to be defined, as well as defining a rotation of members to (1) share the burden of selection, and (2) ensure that different community members can participate in the process, increasing citizen engagement and local governance. A grievance system to respond to beneficiary complaints and ensure a high level of social accountability is desirable. Together these can detect systemic problems to be addressed. These processes require thorough training of and guidance for community committees and an adequate and easily accessible information system for enrollment, transactions, and proper grievance and appeal mechanisms to compensate for the inherent errors and limitations occurring in implementation.

Conclusion

It is hard to summarize both fully and succinctly such a long chapter, spanning diverse targeting methods and with many technical details for each, but this conclusion points out some common threads.

Policy makers must make considered judgments. Chapter 5 discusses the choice of targeting method(s). Even after that choice is made, further judgments are needed in designing the implementation of any given method—about things like the choices over the unit of assistance, the weight put on errors of exclusion versus errors of inclusion, and the emphasis on targeting accuracy versus administrative costs versus incentive effects versus transparency or ownership.

There is no reason to be purist about targeting methods. Many, many countries and programs use multiple methods. Moreover, the line between means testing and HMT is blurry as is that between HMT and PMT. Further, CBT and PMT are increasingly combined.

Data matter. Traditional data, in the sense of government-held administrative data, data from applicant interviews, or community members' knowledge of their neighbors, still dominate in targeting practice. The revolution in people holding and using foundational or functional IDs, especially eIDs, and increased computing power are making it far easier to create integrated or interoperable data systems that lower costs and increase the dynamism of social registries. As data coverage and quality improves, more countries will meet the minimum conditions to move

toward means testing or HMT. Meanwhile, although further work is needed to develop fully and assess its accuracy, big data, such as remote sensing, CDR, and social media data, hold the potential for more frequent and cheaper geographic targeting for various purposes—allocation of program benefits, allocation of program administrative resources, and blending of spatial analysis into formulae for eligibility assessment. Further data advances are on the horizon; how soon and to what extent they can be brought to bear in social assistance programs depends on a series of factors, many of them more about culture and regulation than data science.

Inference matters. There is not a single recipe to guide the modeling in HMT, PMT, or PMTplus, but there is a well-developed body of statistics with applications to targeting that helps guide the modeler, and this chapter has reprised some of the basics. The bottom line for the moment is that there may be a sort of "Goldilocks" range in terms of sophistication—in means testing, capturing most income and verifying some may be enough, and pushing to the extreme may be counterproductive; in PMT modeling, while the added sophistication of quantile regressions or logistic regressions with Lasso selection over principal component analysis or OLS is usually preferable, so far, the complexity of machine learning does not seem to pay off well, at least on the traditional static data used in PMTs. However, machine learning can be used in data preprocessing to create "deep features" within traditional survey data, which may then improve PMT performance regardless of whether traditional regressions or machine learning algorithms are used.

Customization matters. There are a lot of principles that apply and tricks of the trade. But in the end, what makes sense to do must account for specific features of the setting—the goals, design, and budget of the social protection programs; the shape of the welfare distribution; the availability, quality, and details of the data available; national capacities; and political economy.

Good data and inference are important but not sufficient; all targeting mechanisms build on the rest of the delivery system and all require a bridge between the central administration and the potential population. In the more formal systems, building these bridges to bring the local authorities as close as possible to the program implementation activities seems to be largely a matter of delivery systems focused on establishing rules, data systems, and training professional staff. In all, the participation of government staff (locally posted staff of federal agencies or staff at the municipal level) with proper resources and incentives and proper citizen engagement legitimizes the process and helps to improve program outcomes.

The development of good targeting systems takes a diverse skill set. As this chapter has made clear, data/stats nerds clearly have their place on a team. However, as prior chapters have made clear, there is also a need for

the logisticians who plan for good delivery systems; the social workers or community development officers who build bridges between ministries, communities, and individuals; and the advocates, organizers, technocrats, and elected officials who build consensus for good social protection policy.

Targeting methods evolve. Although the suite of methods has the same names as were written about two decades ago, the practice and potential of each is changing as new data, new technology, or new capacities and expectations push them to evolve.

Notes

1. https://www.nytimes.com/2021/03/01/technology/china-national-digital-currency.html
2. See Waze (2018, 2020), https://www.waze.com/wazeforcities/casestudies.
3. https://www.businessinsider.com/how-google-retains-more-than-90-of-market-share-2018-4.
4. The most popular area-level poverty map is that of Fay and Herriot (1979). The Elbers, Lanjouw, and Lanjouw (2003) methodology has since been supplanted in many cases by the empirical best/Bayes approach of Molina and Rao (2010). This chapter focuses on poverty maps developed with big data and does not review the different econometric models currently used in traditional poverty maps. For a good review of the models, see Molina, Corral, and Nguyen (2021). For implementation of the models in Stata, see Corral Rodas, Molina, and Nguyen (2021).
5. Head et al. (2017) use the same transfer learning procedure as Jean et al. (2016) with a different set of countries and an expanded set of human development indicators at the level of household sample survey clusters. Steele et al. (2017) use satellite imagery and CDRs to predict regional poverty and wealth measures. The authors employ hierarchical Bayesian geostatistical models for the prediction task. The models serve as a way of averaging or smoothing across a spatial area while also accounting for uncertainty.
6. Several papers demonstrate a correlation between phone activity (as measured using CDR) and regional economic activity. Toole et al. (2015) use CDR to predict employment. Schmid et al. (2017) predict literacy using CDR. Deville et al. (2014) generate population estimates using CDR, and Blumenstock and Eagle (2012), Dong et al. (2014), and Schmid et al. (2017) use CDR to predict demographic characteristics at the subnational level. Eagle, Macy, and Claxton (2010) explain regional economic rankings using social network diversity as measured using CDR. Frias-Martinez and Virseda (2012) also show that regional economic indicators can be predicted by CDR activity. Mao et al. (2013) and Smith-Clarke, Mashhadi, and Capra (2014) present similar results for Côte d'Ivoire. Pokhriyal and Jacques (2017) predict regional poverty in Senegal using a combination of CDR and satellite imagery. Njuguna and McSharry (2017) also use CDR and satellite imagery to predict the regional Multidimensional Poverty Index in Rwanda using a least absolute shrinkage

and selection operator (Lasso) model. Blondel, Decuyper, and Krings (2015) provide a survey of CDR-based research.

7. The authors implement a deterministic finite automaton to generate 5,088 covariates. They then use the elastic net algorithm to predict the principal component analysis wealth index for each survey respondent.

8. Lain (2018) used data taken from GoogleMaps and Trafi to show that it takes residents of an average Jakarta neighborhood around 40 minutes to reach a regional public hospital using only public transport, but that for some neighborhoods the travel time is almost two hours. Similarly, it costs Jakartans Rp 4,000 (US$0.40) to reach a major hospital at the median, but this rises to Rp 20,000 (US$2) for some neighborhoods. Accessing the three top-ranked senior high schools in each municipal district takes 107 minutes and costs Rp 16,000 (US$16) for some Jakarta residents, despite taking just 35 minutes and Rp 3,500 (US$0.35) for the median neighborhood. Additionally, the times taken to reach the top-ranked high schools are positively correlated with neighborhood-level poverty rates, consistent with the idea that richer households are located in areas that offer better access to good schools, but that do not necessarily offer improved access to other facilities.

9. Facebook data include the number of users, sex, age, reported education type, type of operating system (iOS, Android, or other), expense of phone, and type of connection (for example, 2G, 3G, and so forth).

10. See Coady et al. (2021). The authors estimate the share of means-tested programs among the following types of programs: unemployment, social exclusion, housing, and family and children's programs.

11. See OECD (2013).

12. Urban Institute 2021, quoted in https://fortunly.com/statistics/welfare-statistics/#gref.

13. See Moffitt (2015).

14. Moffitt (2015, 7) explains that the patchwork of means-tested programs in the United States, although it appears as a "crazy-quilt assortment of programs with different structures and recipient groups, rather than following from some single rational design for assistance for the poor of all types," it does reflect the preference of the voters. For example, most programs are in-kind in nature (for medical care, food consumption and nutritional assistance, housing, and early childhood education) and, when cash is provided, it does not cover all low-income households, but certain deserving categories such as workers (Earned Income Tax Credit), the aged, and the disabled (Supplemental Security Income).

15. See https://www.missoc.org/missoc-database/comparative-tables/.

16. Even in advanced economies, administrative data quality is still an issue. To address this issue, in 2014, the United States enacted the Digital Accountability and Transparency Act, which assesses and compares the completeness, timeliness, quality, and accuracy of federal spending data that agencies submit and the implementation and use of data standards. An agency's data quality is considered good if the completeness, timeliness, and accuracy of the information is at least 80 percent. The latest Government Accountability Office report on this topic finds that only 88 percent of the audited agencies achieve this standard

(GAO 2020). Improvements in data quality could be partly tackled by interoperability and integration, but discrepancies in nonsalary income, profits from partnerships, and capital gains can remain.

17. http://mds.gov.br/area-de-imprensa/noticias/2014/setembro/beneficiarios-do
-bolsa-familia-podem-trabalhar-com-carteira-assinada; https://economia.uol.com
.br/guia-de-economia/bolsa-familia-o-que-e-quem-tem-direito-qual-valor.htm;
https://www.youtube.com/watch?v=2jEaQQ01asY.

18. World Bank (2019a).

19. Interoperability is the ability of a system to share information with other systems using common standards.

20. Data integration combines data from different sources and provides users a unified view of these data.

21. Broad-based categorical eligibility is a policy in which households may become categorically eligible for SNAP because they qualify for a noncash TANF or state maintenance of effort funded benefit. Many states implement broad-based categorical eligibility, the programs that confer broad-based categorical eligibility, the asset limit of the TANF/maintenance of effort program, and the gross income limit of the TANF/maintenance of effort program. Broad-based categorical eligibility cannot limit eligibility. Households that are not eligible for the program that confers categorical eligibility may apply for and receive SNAP under regular program rules. Under regular program rules, SNAP households with elderly or disabled members do not need to meet the gross income limit but must meet the net income limit.

22. The threshold formula is set as follows € (189.66 + 132.76 × A + 94.83 × B) where A is the total number of other adults, and B is the total number of children. For example: for a single-family household, the threshold is €189.66 (US$213.10); for a household with two adults and two children, the threshold is €512.08 (US$575.37) (189.66+132.76+94.83 x 2); and for a household with three adults and one child, it is €550.01 (US$617.99) (189.66+132.76 x 2+94.83).

23. The possession of certain kinds of luxury assets such as boats, airplanes, luxury cars, and real estate valued over €150,000 (US$168,540) also acts as a disqualifier filter. This amount is obtained by the following formula: total taxable value may not exceed €90,000 (US$101,124) for the first individual, which is increased by €15,000 (US$16,853.97) for each additional household member, with an overall maximum threshold for each recipient unit of €150,000 (US$168,540).

24. For example: South Africa's Older Persons Grant, Cabo Verde's Minimum Social Pension, and China's Dibao program, which now have more strict verification procedures.

25. Educational level of individuals and school enrollment database.

26. The Social Household Registry is the only registry that collects information on household composition, the assistance unit of the registry. This is self-reported by the head of the household. The legal definition of the household is a group of persons who share the same address, roof, and financial resources. Unlike the United States or the European Union, income taxes in Chile are only individual; hence, information on the family or household is not collected by the tax authorities.

27. The self-employed get a Social Integration Program (PIS) number, which is equivalent to the labor card number. At any time, if the self-employed individual is in formal employment, the PIS number gets converted to a labor card number.

28. About US$500 million as of February 2020.

29. This included 30.3 million social insurance (pensions) beneficiaries, 18.6 million social assistance beneficiaries, and 6.7 million in labor market and activation benefit beneficiaries.

30. After 2018, the Ndhima Ekonomike was switched to PMT.

31. The Ministry of Family and Social Services is in the process of assessing new targeting methods and household needs, particularly those associated with recent economic shocks and increases in poverty.

32. Such data limitations could reduce the accuracy and even feasibility of means testing as well.

33. The Poverty Scorecards developed by Mark Schreiner of Microfinance Risk Management, L.L.C., are a PMT that uses only 10 easy-to-verify pieces of information to help target the benefits of microcredit services. The Poverty Scorecard weights are estimated using a logistic regression, but the main feature is that program staff can compute poverty scores in real time, using paper-based or electronic data collection. However, Diamond et al. (2016) show that using different estimation models with more variables outperforms the nationally calculated simple Poverty Scorecard in terms of bias and variance, highlighting the fundamental trade-off between simplicity of use and accuracy.

34. See Ward et al. (2010) for more details on the targeting precision of the method.

35. According to the Law of August 2020.

36. See World Bank and UNICEF (2020). The vulnerability score in Armenia weights the following household circumstances: the socioeconomic group of each individual in the family, the number of family members with reduced work capacity, area of residence, housing conditions, possession of a vehicle, entrepreneurial activities, conspicuous expenditures (for example, real estate purchases, foreign exchange transactions, electricity, and natural gas), family income, and assessment of the family's living conditions by the social worker.

37. World Bank (2019a).

38. See Areias et al. (forthcoming).

39. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4054530/.

40. An alternative is to produce multiple models, optimized for different program coverage levels. However, this takes significantly more work and can be confusing for policy makers.

41. See del Ninno and Mills (2015) and Brown, Ravallion, and van de Walle (2016).

42. The number of districts and provinces has changed from time to time; this is the number that was pertinent when the work was done.

43. In theory, all possible variable combinations could be tried to determine which model specification performs the best. This approach (see James et al. 2013), which is called "best subset selection," assesses the best model with one variable, two variables, three variables, and up to *p* variables. It reduces selection

among the $2p$ possibilities. This represents $p + 1$ possible models by selecting the best model for each exactly k variable, where $k = 1, 2, …, p$, and uses cross-validated prediction errors to assess between the best models for each k. The residual sum of squares or R^2 will always indicate that the model with all the variables performs the best, but this indicates that it describes the training data the best and not necessarily how well it will perform out of sample. The problem with "best subset" is that it is very computationally intensive for any significant number of potential variables. Forward stepwise begins with a model with no variables and assesses all the models with a single variable included to see which improves cross-validated prediction error the most over the null model. The variable that led to the best single-variable model is retained. Then all possible combinations of that variable with one other variable are assessed. The variables from the two-variable model with the best cross-validated prediction error are then kept as the basis for a three-variable model and so forth until no further improvements are obtained by adding new variables. Backward stepwise begins with a model with all the variables included and subtracts one variable at a time, again looking for the model that has the best cross-validated prediction errors, and continues subtracting variables until no further improvements are made.

44. These are called penalized regressions or shrinkage methods; see Areias et al. (forthcoming) for a full review.

45. Lasso is used as a variable selection device and a less biased post-Lasso estimator is used to calculate the final PMT weights. OLS is then used to regress log consumption on the Lasso selected variables and derive the final scoring coefficients.

46. However, there are differences between explanatory and predictive modeling, which Shmueli (2010) highlights. First, Shmueli indicates that predictive modeling has higher predictive accuracy than explanatory statistical models. Second, Shmueli highlights that predictive models: (1) aim to look for association between the X (covariates) and Y (dependent variable); (2) do not have a requirement for direct interpretability in terms of the relationship between X and Y; (3) have a forward-looking approach instead of testing existing hypotheses; and (4) reduce at once the combination of bias (result of misspecification of the model) and estimation variance (result of using a sample). Addressing these points in predictive models translates into a different approach for selecting the covariates. The main criteria for selecting the set of covariates are the quality of the association between them and the dependent variable, as well as preexisting knowledge of correlation/association that does not necessarily come from the data set but from other studies or local knowledge. This procedure is quite different in explanatory models, where researchers (1) only keep significant variables in the model, (2) must address multicollinearity, (3) must have clear/independent control variables, and (4) must minimize endogeneity to address causality.

47. See Elbers, Lanjouw, and Lanjouw (2003).

48. Few expenditure surveys collect such variables, which often have very good discretionary power and can be verifiable as part of the presentation of required documents during enrollment.

49. Sample size issues can be addressed by pooling multiple rounds of survey data if they are frequent (for example, annually). In Indonesia, the greatest improvement in predictive accuracy came from moving from one year of data used for 65 provincial urban-rural models to using three years of pooled data, which allowed the reliable estimation of district-level models (around 515 models).

50. See Stock and Watson (2003).

51. Model selection in predictive modeling is not based on the explanatory power assessed using metrics computed as R^2-type values and the statistical significance of overall f-type statistics. The researcher can retain covariates that are statistically insignificant if the variables are important for the prediction. Predictive power for predictive models is measured by their capacity to predict the event using new data (Geisser 1975; Stone 1974) or carefully using the same data. Usually, researchers focus on extracting a holdout (subsample from the same data) or pseudo-samples. In the targeting context, beyond measuring whether the average prediction and errors are acceptable overall, researchers must analyze the predictive power of the model for certain marginalized groups or groups that must be targeted by the method. For example, good predictive power for the average income or poverty levels in a region X does not guarantee that the same power would generate acceptable errors for households with elderly living alone, small households, or female-headed households. PMT uses statistical models for prediction. This means significant differences in use compared with statistical models for identifying causal relationships. Causal models relate the dependent variable (the variable being predicted) to a set of independent (or explanatory) variables. They assume that the explanatory covariates are unrelated to each other and the dependent variable does not have any reverse causal relationship to any of the covariates. With PMT, inference is not causal but about association. Hence, the strong underlying assumptions needed to determine causality are nonexistent or incorporated in a less formal way. Consequently, the best PMT model is not the one with high explanatory power or R^2, but the one with high predictive power, which can be quite different.

52. SISBEN I started its data collection in 1995, SISBEN II in 2005, SISBEN III in 2011, and SISBEN IV in 2017. SISBEN III was the main database used to allocate services and benefits until March 2020, when SISBEN IV data started to be used.

53. See Hillis et al. (2013).

54. The pilot has been in full operation since the end of 2009. The pilot program was planned and designed in 2008. Targeting, enrollment, management information systems development, and all other preparations were developed in 2009. The first cash transfer payment was delivered to beneficiaries in November–December 2009. The first phase of enrollment covered almost 2,500 households. In August 2011, it reached 4,998 households in 40 villages in three districts in two provinces: Chamwino in Dodoma, and Bagamoyo and Kibaha in Pwani. The targeting method combined geographic targeting, CBT, and PMT.

55. See Bourguignon, Ferreira, and Leite (2008) and Datt and Ravallion (1991).

56. This is a common test used in poverty map methodology for comparing household survey and census data distributions. See Elbers, Lanjouw, and Lanjouw (2003).

57. Most of the household surveys combine stratified sampling with cluster sampling, and without replacement.

58. For example, the predictive power of any inference test drops quickly as intra-cluster correlations increase.

59. A third source of prediction error is error variance, which is irreducible noise that cannot be eliminated through modeling.

60. The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a data set with replacement to create many simulated samples. The jackknife method (or leave-one-out method) is an alternative resampling method to the bootstrap, based on sequentially deleting one observation from the data set to estimate the precision of the estimator.

61. See Areias et al. (forthcoming) for discussion of a special case of k-fold validation called leave-one-out cross validation (LOOCV) where k is set to n, the number of observations in the data.

62. The challenge is that with cross-sectional data, household welfare before and after a shock is not observed. Instead, the impact of a shock on changes in consumption is inferred from differences in consumption of otherwise observationally equivalent households.

63. The Listahanan or National Household Targeting System for Poverty Reduction is an information management system that identifies who and where the poor are in the Philippines.

64. The models used are (1) current: the current formula (the default); (2) recalibrate: recalibrate the coefficients of the current model with the Household Socio-Economic Survey 2018 data; (3) new: regression models with more and/or different variables than the current model; (4) stepwise: use stepwise regressions to streamline variable selection in the new model; (5) Lasso: use the least absolute shrinkage and selection operator (Lasso) algorithm for variable selection in the new model; (6) Lasso-int: use the Lasso algorithm for variable selection in the new model and extend it to include variable interactions; (7) random forest: use the random forest algorithm on the variables for the new model to predict welfare (this algorithm is nonlinear and does not result in simple coefficients for each of the variables); and (8) Nnet: use the neural network algorithm on the variables for the new model to predict welfare (this algorithm is also nonlinear and does not result in simple coefficients for each of the variables).

65. The machine learning models used are Ridge, Lasso, elastic net, random forest, gradient boosting, and two blended models.

66. In the Costa Rican Household Poverty Level Kaggle Competition, 616 teams tried to build the most accurate poverty classifier; all the top performers with public code used feature engineering and tree-based models (such as XGB and LightGBM). See https://www.kaggle.com/c/costa-rican-household-poverty-prediction/overview.

67. Quantile regression with cross-validated quantile.

68. Targeting measures are discussed in detail in chapter 7. For readers already familiar with inclusion and exclusion errors, Precision is 1 – Inclusion Error (or the

percentage of predicted eligible who are truly *eligible*; Recall is 1 – Exclusion Error (or the percentage of those truly eligible who are *correctly* predicted so). F_1 is then the harmonic mean of Precision and Recall, or 2 * ([Recall * Precision] / [Recall + Precision]). In this sense, it is used as a balanced average of measures analogous to inclusion and exclusion errors. F_2 is the same as F_1 but with twice the weight on Recall (that is, exclusion error carries more weight). Matthews correlation coefficient (MCC) is the geometric mean of Informedness and Markedness, where Informedness is Recall + Specificity – 1, where Specificity is the percentage of truly ineligible households correctly predicted so (or TN / [TN + FP]) and Markedness is Precision + Inverse Precision – 1, where Inverse Precision is the percentage of those predicted not eligible who are truly not eligible (or TN / [FN + TN]). Effectively, MCC is a scale-invariant balance of inclusion and exclusion errors.

69. Areias et al. (forthcoming); McBride and Nichols (2016); Ohlenburg (2020b); Shrestha (2020).
70. Gradient boosted quantile regression trees tend to outperform most other algorithms under both implementation modalities (line and quota), but the traditional logistic regression with a simple Lasso variable selection model also proves robustly useful.
71. *k*-nearest neighbors.
72. The choice between gradient boosted quantile regression trees and logistic regression with a simple Lasso variable may depend on other considerations. The former allows for more predictors than observations (which is quite feasible with big data), meaning a greater range of models can be explored, whereas the latter does not. Logistic models also require significant data preprocessing and are not robust to predictor noise, while boosted models do not require preprocessing and are robust. Logistic models are easier to interpret than boosted models, do not require optimization or choice of tuning parameters, and are computationally much less intensive.
73. Areias et al. (forthcoming) implement both a threshold approach, where only households with scores below a certain threshold are deemed eligible, regardless of how many there are, and a quota approach, where the scores are only used to rank households and a fixed quota (tied to, say, the official poverty rate or the program budget) is deemed eligible.
74. Personal communication.
75. A similar approach is being implemented in Panama (https://www.gaceta oficial.gob.pa/pdfTemp/29163/GacetaNo_29163_20201126.pdf) and Honduras (https://www.iadb.org/es/noticias/honduras-mejora-las-condiciones-de-vida -de-los-hogares-mas-pobres-con-apoyo-del-bid).
76. See Steele et al. (2017) and https://blogs.worldbank.org/opendata/using-big -data-and-machine-learning-locate-poor-nigeria.
77. As COVID-19 cases increase, the government of Bangladesh with the support of Yale Y-Rise is working with machine learning modelers to extrapolate trends in mobile usage data to identify those who need to be targeted with cash support.
78. Blumenstock, Cadamuro, and On (2015); Head et al. (2017); and Jean et al. (2016) use Demographic and Health Survey data to train their models. The targeting performance for all methods in Aiken et al. (2021) improve when

moving from a PMT-based dependent variable to a true consumption variable, suggesting that the use of proxies for training models introduces additional noise (although all methods face this issue, not just ones based on big data).

79. See, for example, global mobility trends over COVID-19: https://ourworldindata .org/covid-google-mobility-trends.

80. See, for example, Beegle et al. (2012) on the impact of truncated consumption modules.

81. See Montjoye, Gambs, and Blondel (2018) and Zhang, Chen, and Zhong (2016).

82. Camacho and Conover (2011) find that in Colombia's traditional PMT, there is anecdotal evidence of people moving or hiding their assets or borrowing and lending children, and there is evidence of manipulation of scores.

83. The Participatory Wealth Ranking begins with a communitywide meeting convened by the facilitation team. After discussing the meaning and understanding of poverty in the local context, the people draw a map of all the households in the village and fill a card with the name of each household. Three reference groups are then formed in each ranking section, that is, the hamlet. McCord (2013) highlights that the Participatory Wealth Ranking literature shows that it delivers a robust and broadly noncontentious ranking based on a community's own understanding of poverty considered from multiple dimensions.

84. The number of Ubudehe categories evolved over time. At the start of Ubudehe categorization system in 2002, there were six categories. These were revised to four categories in 2014. Currently, the government is in the process of introducing a five-category system. In the current four Ubudehe categories, the first category is designated for the poorest people in society, while the fourth category is for the wealthiest members of society. More specifically: category 1: very poor and vulnerable citizens who are homeless and unable to feed themselves without assistance; category 2: citizens who are able to afford some form of rented or low-class owned accommodation, but who are not gainfully employed and can only afford to eat once or twice a day; category 3: citizens who are gainfully employed or even employers of labor (this category includes small farmers who have moved beyond subsistence farming, or owners of small and medium-scale enterprises); and category 4: citizens classified under this category are chief executive officers of large businesses, employees who have full-time employment with organizations, industries or companies, government employees, owners of lockdown shops or markets, and owners of commercial transport or trucks (Government of Rwanda 2015; MINALOC 2015).

85. The index was using factor analysis on a set of variables such as region, type of dwelling, house owner, poverty level, consumption expenditure, type of residency, current dwelling value, water source, light source, amount paid for electricity in the past four weeks, primary source of cooking fuel, and type of toilet in an aggregated form using factor analysis.

86. i2i Dime and World Bank (2017).

87. Area Executive Committees are mostly government frontline staff, also known as extension workers at the community/traditional authority level, for example, community development assistants; health surveillance assistants; and

agricultural extension workers. Sometimes nongovernmental organizations also dispatch their own frontline staff to be a member of the Area Executive Committee.

88. In Mali's Jigisemejiri program, the committees are paid for their attendance, which represents a significant operational cost, but program administrators use committees for other functions.

References

Aiken, Emily, Suzanne Bellue, Dean Karlan, Christopher R. Udry, and Joshua Blumenstock. 2021. "Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance." NBER Working Paper 29070, National Bureau of Economic Research, Cambridge, MA.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2019. "Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia." *AEA Papers and Proceedings* 109: 334–39. https://doi.org/10.1257/pandp.20191047.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 102 (4): 1206–40.

Al-Iryani, Lamis, Alain de Janvry, and Elisabeth Sadoulet. 2015. "The Yemen Social Fund for Development: An Effective Community-Based Approach amid Political Instability." *International Peacekeeping* 22 (4): 321–36.

Areias, Ana, Niall Keleher, Laura Ralston, and Matthew Wai-Poi. Forthcoming. "Big Data, Machine Learning and Prediction in the Targeting of Social Protection Programs." World Bank, Washington, DC.

Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2 (1): 1–30.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson. 2020. "Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi." *Journal of Public Economics* 185 (104047).

Baum, Tinatin, Anastasia Mshvidobadze, and Josefina Posadas. 2016. *Continuous Improvement: Strengthening Georgia's Targeted Social Assistance Program*. Directions in Development Human Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/24812.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson. 2012. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics* 98 (1): 3–18. https://ideas.repec.org/a/eee/deveco/v98y2012i1p3-18.html.

Ben-Avie, Jochai, and Udbhav Tiwari. 2019. "India's New Data Protection Bill: Strong on Companies, Step Backward on Government Surveillance." *Blog: Open Policy & Advocacy*, December 10. https://blog.mozilla.org/netpolicy/2019/12/10/indias-new-data-protection-bill-strong-on-companies-weak-on-gov.

Bergeron, Gilles, Saul Sutkover Morris, and Juan Manuel Medina Banegas. 1998. "How Reliable Are Group Informant Ratings? A Test of Food Security Rating in Honduras." *World Development* 26 (10): 1893–902.

Björkegren, Daniel, Joshua E. Blumenstock, and Samsun Knight. 2020. "Manipulation-Proof Machine Learning." ArXiv:2004.03865 [Cs, Econ], April. http://arxiv.org/abs/2004.03865.

Blondel, Vincent D., Adeline Decuyper, and Gautier Krings. 2015. "A Survey of Results on Mobile Phone Datasets Analysis." *EPJ Data Science* 4 (1): 10.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350: 1073. https://doi.org/10.1126/science.aac4420.

Blumenstock, Joshua, and Nathan Eagle. 2010. "Mobile Divides: Gender, Socioeconomic Status, and Mobile Phone Use in Rwanda." *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development* 6: 1–6. ICTD 2010, New York.

Blumenstock, Joshua, and Nathan Eagle. 2012. "Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda." *Information Technologies & International Development* 8 (2): 1–16.

Boko, Joachim, Luis Alberro, Maimouna Gueye, Audrey Ariss, and Tina George. 2020. "Technology & Innovations for Africa's Human Capital in the Face of the Covid-19 Crisis: Technical Note on Togo's Novissi Program." World Bank, Washington, DC.

Bourguignon, François, Francisco H. G. Ferreira, and Phillippe G. Leite. 2008. "Beyond Oaxaca–Blinder: Accounting for Differences in Household Income Distributions." *Journal of Economic Inequality* 6 (2): 117–48.

Brown, Caitlin, Martin Ravallion, and Dominique van de Walle. 2016. "A Poor Means Test? Econometric Targeting in Africa." NBER Working Paper No. 22919, National Bureau of Economic Research, Cambridge, MA.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15.

Camacho, Adriana, and Emily Conover. 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy*, 3 (2): 41–65.

Carraro, Ludovico. 2014. "Ajutorul Social Assessment Report." World Bank, Washington, DC.

Carrillo, Alfredo, Luis F. Cantú, Luis Tejerina, and Alejandro Noriega. 2021. "Individual Explanations in Machine Learning Models: A Case Study on Poverty Estimation." arXiv:2104.04148v2 [cs.LG]. https://arxiv.org/abs/2104.04148.

Chambers, Robert. 1994. "Participatory Rural Appraisal (PRA): Challenges, Potentials and Paradigm." *World Development* 22 (10): 1437–54. ISSN 0305-750X, https://doi.org/10.1016/0305-750X(94)90030-2.

Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. 2021. "Micro-Estimates of Wealth for all Low- and Middle-Income Countries." Working Paper, Cornell University, Ithaca, NY.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/14902.

Coady, David, Samir Jahan, Riki Matsumoto, and Baoping Shang. 2021. "GMI Schemes in Europe: Landscape and Design." IMF Working Paper WP/21/179, International Monetary Fund, Washington, DC.

Conning, Jonathan, and Michael Kevane. 2002. "Community-Based Targeting Mechanisms for Social Safety Nets: A Critical Review." *World Development* 30 (3): 375–94.

Cook, Benjamin Lé, and Willard G. Manning. 2013. "Thinking beyond the Mean: A Practical Guide for Using Quantile Regression Methods for Health Services Research." *Shanghai Archives of Psychiatry* 25 (1): 55–59.

Corral Rodas, Paul, Isabel Molina, and Minh Nguyen. 2021. "Pull Your Small Area Estimates Up by the Bootstraps." *Journal of Statistical Computation and Simulation*. https://doi.org/10.1080/00949655.2021.1926460.

Council of Ministers. 2005. "Decision No. 787/2005 on Determining the Criteria and Procedures of Economic Assistance Levels." Council of Ministers, Tirana, Albania.

CRES (Centre des Recherches et des études Sociales) and World Bank. 2018. "Evaluation des données administratives Relative au Revenu des Ménages en Tunisie." World Bank, Washington, DC.

Datt, Gaurav, and Martin Ravallion. 1991. "Regional Disparities, Targeting, and Poverty in India." In *Including the Poor*, edited by Michael Lipton and Jacques van der Gaag, 91–114. Washington, DC: World Bank.

de Montjoye, Yves-Alexandre, Sébastien Gambs, Vincent Blondel, Geoffrey Canright, et al. 2018. "On the Privacy-Conscientious Use of Mobile Phone Data." *Scientific Data* 5: 180286. https://doi.org/10.1038/sdata.2018.286.

del Ninno, Carlo, and Bradford Mills. 2015. *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*. Washington, DC: World Bank. https://doi .org/10.1596/978-1-4648-0435-9.

Deville, Pierre, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem. 2014. "Dynamic Population Mapping Using Mobile Phone Data." *Proceedings of the National Academy of Sciences* 111 (45): 15888–893.

Diamond, Alexis J., M. Gill, M.A. Rebolledo Dellepiane, Skoufias, E., Vinha, K., Xu, Y. 2016. "Estimating Poverty Rates in Target Populations: An Assessment of the Simple Poverty Scorecard and Alternative Approaches." World Bank Policy Research Working Paper No. 7793, https://ssrn.com/abstract =2836540.

Dong, Yuxiao, Yang Yang, Jie Tang, and Nitesh V. Chawla. 2014. "Inferring User Demographics and Social Strategies in Mobile Social Networks." In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 15–24. New York.

Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. "Network Diversity and Economic Development." *Science* 328 (5981): 1029–31.

Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. "Micro–Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–64.

Engstrom, Ryan, Jonathan Samuel Hersh, and David Newhouse. 2021. "Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being." *The World Bank Economic Review,* lhab015, https://doi.org/10.1093 /wber/lhab015.

Fatehkia, Masoomali, Isabelle Tingzon, Ardie Orden, Stephanie Sy, Vedran Sekara, Manuel Garcia-Herranz, and Ingmar Weber. 2020. "Mapping Socioeconomic Indicators Using Social Media Advertising Data." EPJ Data Science 9: 22. https://doi.org/10.1140/epjds/s13688-020-00235-w.

Fay, Robert E., and Roger A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 85: 398–409. https://doi.org/10.1080/01621459.1979.10482505.

Feulefack, Joseph, and Manfred Zeller. 2005. "How Accurate Is Participatory Wealth Ranking (PWR) in Targeting the Poor? A Case Study from Bangladesh." Conference on International Agricultural Research for Development, Stuttgart-Hohenheim, Germany, October 11–13.

Filmer, Deon, and Lant H. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India." *Demography* 38 (1): 115–32 (accessed August 19, 2020), www.jstor.org/stable/3088292.

Fortmann-Roe, Scott. 2012. "Understanding the Bias-Variance Tradeoff." https://scott.fortmann-roe.com/docs/BiasVariance.html.

French Republic. 2016. "LOI No. 2016-1321 of October 7, 2016 for a Digital Republic." 2016-1321. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746?r=LhMomAa3M7.

Frias-Martinez, Vanessa, and Jesus Virseda. 2012. "On the Relationship between Socio-Economic Factors and Cell Phone Usage." *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*, 76–84. New York.

GAO (Government Accountability Office). 2020, "Audit of the Data Act." GAO, Washington, DC. https://www.gao.gov/assets/gao-20-540.pdf.

Geisser, Seymour. 1975. "The Predictive Sample Reuse Method with Applications." *Journal of the American Statistical Association* 70 (350): 320–28.

Gelb, Alan, and Anna Diofasi Metz. 2018. "Identification Revolution: Can Digital ID Be Harnessed for Development?" Brief, Center for Global Development, Washington, DC.

Golan, Jennifer, Terry Sicular, and Nithin Umapathi. 2014. "Any Guarantees? China's Rural Minimum Living Standard Guarantee Program." Social Protection and Labor Discussion Paper No. 1423, World Bank Group, Washington, DC. https://openknowledge.worldbank.org/handle/10986/1997.

Government of Kyrgyz Republic. 2018. "Resolution 307 on the Implementation of the Law on State Benefits in the Kyrgyz Republic." Bishkek, Kyrgyz Republic.

Greene, William H. 2011. *Econometric Analysis*, seventh ed. Pearson. ISBN-13: 978-0131395381, ISBN-10: 0131395386.

Günther, Isabel, and Kenneth Harttgen. 2009. "Estimating Household Vulnerability to Idiosyncratic and Covariate Shocks: A Novel Method Applied in Madagascar." *World Development* 37 (7): 1222–34.

Handa, Sudhanshu, Carolyn Huang, Nicola Hypher, Clarissa Teixeira, Fabio V. Soares, and Benjamin Davis. 2012. "Targeting Effectiveness of Social Cash Transfer Programmes in Three African Countries." *Journal of Development Effectiveness* 4 (1): 78–108.

Hardt, Moritz, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. 2015. "Strategic Classification." arXiv:1506.06980 [cs.LG]. https://arxiv.org/abs/1506.06980.

Head, Andrew, Mélanie Manguin, Nhat Tran, and Joshua E. Blumenstock. 2017. "Can Human Development Be Measured with Satellite Imagery?" *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*. Article 8. https://doi.org/10.1145/3136560.3136576.

Hill, Kashmir. 2020. "Wrongfully Accused by an Algorithm." *New York Times*, August 3, Technology Section. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

Hill, Ruth V., and Catherine Porter. 2017. "Vulnerability to Drought and Food Price Shocks: Evidence from Ethiopia." *World Development* 96: 65–77.

Hillis, Samira, Phillippe Leite, Laura McDonald, and Lillian Frost. 2013. "Designing an Advanced Cash Transfer Program in the Palestinian Territories." MENA Knowledge and Learning Quick Notes No. 98, World Bank, Washington, DC.

Himmelstine, Carmen Leon, and Anna McCord. 2012. "The Use of Community-Based Targeting in Low- and Middle-Income Countries." Annotated Bibliography, Overseas Development Institute, London.

Holzmann, Penny, Tanya Boudreau, Julius Holt, Mark Lawrence, and Michael O'Donnell. 2008. *The Household Economy Approach: A Guide for Programme Planners and Policy-Makers*. New York: Save the Children. https://resourcecentre.savethechildren.net/node/13676/pdf/hea_guide.pdf.

Honorati, Maddalena, Roberto Claudio Sormani, and Ludovico Carraro. 2020. "Assessing the Targeting System in Georgia: Proposed Reform Options." Social Protection and Jobs Discussion Paper No. 2005, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34358. License: CC BY 3.0 IGO.

i2i Dime and the World Bank. 2017. "Field Experiments on Targeting Labor Intensive Public Works in Eastern Congo." World Bank, Washington, DC. https://www.jobsanddevelopment.org/wp-content/uploads/2019/04/Congo_LIPW.pdf.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer Verlag.

Jean, Neal, Marshall Burke, Michael Xie, W. Michael Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353 (6301): 790–94.

Kebede, Bereket. 2009. "Community Wealth Ranking and Household Surveys: An Integrative Approach." *Journal of Development Studies* 45: 1731–46. https://doi.org/10.1080/00220380902935832.

Kidd, Stephen. 2011. "Targeting the Poorest: An assessment of the Proxy Means Test Methodology." Australian AID, Canberra, Australia. https://doi.org/10.13140/2.1.3567.6488.

Kidd, Stephen, Bjorn Gelders, and Diloá Bailey-Athias. 2017. "Exclusion by Design: The Effectiveness of the Proxy Means Test." ESS Working Paper No. 56, International Labour Office, Development Pathways, Sidcup, UK.

Koenker, Roger, and Gilbert Bassett, Jr. 1978. "Regression Quantiles." *Econometrica* 46 (1): 33–50.

Kuhn, Max, and Kjell Johnson. 2018. *Applied Predictive Modeling*. New York: Springer.

Lain, Jonathan. 2018. "Multi-Dimensional Urban Poverty in Indonesia." World Bank, Washington, DC.

Lange, Simon, Stephan Klasen, Grace Hadiwidjaja, Juul Pinxten, and Bagus Wirapati. 2016. "Evaluation of PMT-Based Targeting in Indonesia." Project Report, World Bank, Washington, DC.

Leite, Phillippe. 2014. "Effective Targeting for the Poor and Vulnerable." Social Protection & Labor Technical Note No. 6, World Bank, Washington, DC.

Leite, Phillippe 2015. "Short- and Long-Term Targeting in Kenya." In *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*, edited by Carlo del Ninno and Bradford Mills. Washington, DC: World Bank. https://doi .org/10.1596/978-1-4648-0435-9.

Lindert, Kathy, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Nishikawa Chávez. 2020. *Sourcebook on the Foundations of Social Protection Delivery Systems*. Washington, DC: World Bank.

Lubarsky, Boris. 2017. "Re-Identification of 'Anonymized' Data." Georgetown Law, Technology Review, Washington, DC. https://georgetownlawtechreview.org /re-identification-of-anonymized-data/GLTR-04-2017/.

Mao, Huina, Xin Shuai, Yong-Yeol Ahn, and Johan Bollen. 2013. "Mobile Communications Reveal the Regional Economy in Côte d'Ivoire." Paper presented at NetMob 2013, Cambridge, MA, May 1–3.

Masaki, Takaaki, David Locke Newhouse, Ani Rudra Silwal, Adane Bedada, and Ryan Engstrom. 2020. "Small Area Estimation of Non-Monetary Poverty with Geospatial Data." Policy Research Working Paper 9383, World Bank, Washington, DC. https://ideas.repec.org/p/wbk/wbrwps/9383.html.

McBride, Linden, and Austin Nichols. 2016. "Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning." *World Bank Economic Review* 32 (3): 531–50.

McCord, Anna. 2013. "Community-Based Targeting in the Social Protection Sector." Working Paper No. 514, Overseas Development Institute, London. https:// medium.com/center-for-effective-global-action/using-mobile-phone-and -satellite-data-to-target-emergency-cash-transfers-f0651b2c1f3f.

MINALOC (Ministry of Local Government). 2015. "Revised Ubudehe Categories." MINALOC, Kigali, Rwanda. www.minaloc.gov.rw.

MISSOC (Mutual Information System on Social Protection). 2020. "Comparative Tables." MISSOC, European Commission, Brussels, Belgium. https://www .missoc.org/missoc-database/comparative-tables/.

MoARD (Ministry of Agriculture and Rural Development). 2006. "Productive Safety Net Programme, Programme Implementation Manual." MoARD, Addis Ababa, Ethiopia.

Moffitt, Robert A. (ed.). 2015. *Economics of Means-Tested Transfer Programs in the United States*. Cambridge, MA: National Bureau of Economic Research. https://www .nber.org/books-and-chapters/economics-means-tested-transfer-programs -united-states-volume-1.

Molina, Isabel, Paul Corral, and Minh Nguyen. 2021. "Model-Based Methods for Poverty Mapping: A Review." Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette, Luxembourg.

Molina, Isabel, and J. N. K. Rao. 2010. "Small Area Estimation of Poverty Indicators." *Canadian Journal of Statistics* / La Revue Canadienne De Statistique 38 (3): 369–85 (accessed June 15, 2021). http://www.jstor.org/stable/27896031.

Nizeyimana, Pacifique, Kee-Won Lee, and Songyong Sim. 2018. "A Study on the Classification of Households in Rwanda Based on Factor Scores." *Journal of the Korean Data & Information Science Society* 29 (2): 547–55.

Njuguna, Christopher, and Patrick McSharry. 2017. "Constructing Spatiotemporal Poverty Indices from Big Data." *Journal of Business Research* 70: 318–27.

Noriega-Campero, Alejandro, Bernardo Garcia-Bulle, Luis Fernando Cantu, Michiel A. Bakker, Luis Tejerina, and Alex Pentland. 2020. "Algorithmic Targeting of Social Policies: Fairness, Accuracy, and Distributed Governance." Conference on Fairness, Accountability, and Transparency (FAT* '20), January 27–30, Barcelona, Spain. ACM, New York. https://doi.org/10.1145/3351095.3375784.

OECD (Organisation for Economic Co-operation and Development). 2013. *Greece: Reform of Social Welfare Programs*. OECD Public Governance Reviews. Paris: OECD.

OECD (Organisation for Economic Co-operation and Development). 2018. *Social Protection System Review of Kyrgyzstan*. OECD Development Pathways. Paris: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). 2019. "Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies." Paris: OECD Publishing. https://doi.org/10.1787/276aaca8-en.

Ohlenburg, Tim. 2020a. "AI in Social Protection: Exploring Opportunities and Mitigating Risks." GIZ, Bonn.

Ohlenburg, Tim. 2020b. "Technical Report: Machine Learning for Proxy Means Testing in Indonesia." World Bank, Washington, DC.

Ortakaya, Ahmet Fatih. 2020. "COVID-19 Social Support Program in Turkey." World Bank, Washington, DC.

Ortakaya, Ahmet Faith. 2021. "BMM Framework in Turkey's Integrated Social Assistance System (ISAS)." Presentation at the SPJ Core Courses, May 15–18. Kuwait.

Paes de Barros, Ricardo, Diana Coutinho, Samuel Franco, and Rosane Mendonça. 2016. "Desenho de um Sistema para Aferição dos Valores dos Benefícions do Programa Renda Melhor." Secretaria de Estado de Assistência Social e Direitos Humanos – SEASDH Governo do Estado do Rio de Janeiro.

Pokhriyal, Neeti, and Damien Christophe Jacques. 2017. "Combining Disparate Data Sources for Improved Poverty Prediction and Mapping." *PNAS* 114 (46): E9783–E9792.

Pople, Ashley, Ruth Hill, Stefan Dercon, and Ben Brunckhorst. 2021. "Anticipatory Cash Transfers in Climate Disaster Response." Working Paper 6, Centre for Disaster Protection, London.

Premand, Patrick, and Pascale Schnitzer. 2018. "Efficiency, Legitimacy and Impacts of Targeting Methods: Evidence from an Experiment in Niger." Policy Research Working Paper 8412, World Bank, Washington, DC.

Roberts, Mark, Frederico Gil Sander, and Sailesh Tiwari. 2019. *Time to ACT: Realizing Indonesia's Urban Potential*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1389-4.

Robertson, Laura, Phyllis Mushati, Jeffrey W. Eaton, Lovemore Dumba, et. al. 2013. "Effects of Unconditional and Conditional Cash Transfers on Child Health and Development in Zimbabwe: A Cluster-Randomised Trial." *Lancet* 381 (9874): 1283–92. https://doi.org/10.1016/S0140-6736(12)62168-0.

Rosenbaum, Dorothy. 2019. "SNAP's 'Broad-Based Categorical Eligibility' Supports Working Families and Those Saving for the Future." Center on Budget and Policy Priorities, Washington, DC.

Schmid, Timo, Fabian Bruckschen, Nicola Salvati, and Till Zbiranski. 2017. "Constructing Sociodemographic Indicators for National Statistical Institutes by Using Mobile Phone Data: Estimating Literacy Rates in Senegal." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180 (4): 1163–90.

Schnitzer, Pascale. 2019. "How to Target Households in Adaptive Social Protection Systems? Evidence from Humanitarian and Development Approaches in Niger." *Journal of Development Studies* 55 (sup1): 75–90.

Scott Morton, Fiona, and David Dinielli. 2020. "Roadmap for a Digital Advertising Monopolization Case against Google." Omidyar Network. https://omidyar.com/roadmap-for-a-monopolization-case-against-google-regarding-the-search-market/.

Sepulveda Carmona, Magdalena. 2018. "Is Biometric Technology in Social Protection Programmes Illegal or Arbitrary? An Analysis of Privacy and Data Protection." Extension of Social Security Working Paper 59, International Labour Organization, Geneva. https://www.researchgate.net/publication/325909014_Is_biometric_technology_in_social_protection_programmes_illegal_or_arbitrary_An_analysis_of_privacy_and_data_protection.

Sepulveda Carmona, Magdalena, and Carly Nyst. 2012. "The Human Rights Approach to Social Protection." Ministry of Foreign Affairs, Helsinki, Finland. https://ssrn.com/abstract=2114384.

Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3): 289–310. https://doi.org/10.1214/10-STS330.

Shrestha, Mahesh. 2020. "Improving the Integrated Household Database System for More Effective Social Assistance in Mongolia." World Bank, Washington, DC.

Silva-Leander, Sebastian, and Fred Merttens. 2016. "HSNP: Assessment of Programme Targeting." Oxford Policy Management, Oxford, UK. https://www.opml.co.uk/files/Publications/a0013-evaluation-kenya-hunger-safety-net-programme/assessment-of-hsnp2-targeting.pdf?noredirect=1.

Skoufias, Emmanuel, and Javier Báez. 2021. "Operationalizing Vulnerability to Poverty." Working Paper, World Bank, Washington, DC.

Skoufias, Emannuel, Yasuhiro Kawasoe, Eric Strobl, and Pablo Acosta. 2019a. "Identifying the Vulnerable to Poverty from Natural Disasters: The Case of Typhoons in the Philippines." Policy Research Working Paper 8857, World Bank, Washington, DC.

Skoufias, Emannuel, Katja Vinha, Alvina Erman, and Stephane Hallegatte. 2019b. "Identifying the Vulnerable to Poverty in Africa." World Bank, Washington, DC.

Smith IV, Jack. 2016. "Crime-Prediction Tool May Be Reinforcing Discriminatory Policing." *Business Insider*, October 10. https://www.businessinsider.com/predictive-policing-discriminatory-police-crime-2016-10?r=UK.

Smith-Clarke, Christopher, Afra Mashhadi, and Licia Capra. 2014. "Poverty on the Cheap: Estimating Poverty Maps Using Aggregated Mobile Communication Networks." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 511–20. CHI 2014, New York: ACM.

Steele, Jessica E., Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M. Iqbal, Khandakar N. Hadiuzzaman, Xin Lu, Erik Wetter, Andrew J. Tatem, and Linus Bengtsson. 2017. "Mapping Poverty Using Mobile Phone and Satellite Data." *Journal of the Royal Society Interface* 14 (127): 20160690.

Stock, James H., and Mark W. Watson. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature* XLI: 788–829.

Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions (with Discussion)." *Journal of the Royal Statistical Society* Ser. B 39: 111–47. http://www.jstor.org/stable/2984809.

Tatman, Rachael. 2017. "Gender and Dialect Bias in YouTube's Automatic Captions." *Proceedings of the First Workshop on Ethics in Natural Language Processing*, 53–59, Valencia, Spain, April 4.

Tesliuc, Emil, Phillippe Leite, and Katerina Petrina. 2009. "Improving Targeting of Social Assistance Programs in the Ukraine." World Bank, Washington, DC.

Tesliuc, Emil, Lucian Pop, Margaret Grosh, and Ruslan Yemtsov. 2014. *Income Support for the Poorest: A Review of the Experience in Eastern Europe and Central Asia*. Directions in Development Human Development. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/18886.

Toole, Jameson L., Yu-Ru Lin, Erich Muehlegger, Daniel Shoag, Marta C. González, and David Lazer. 2015. "Tracking Employment Shocks Using Mobile Phone Data." *Journal of the Royal Society Interface* 12 (107): 20150185.

Turkey MFSP and World Bank. 2016. "Turkey's Integrated Social Assistance System." World Bank, Washington, DC.

Van Domelen, Julie, and Sarah Coll-Black. 2012. *Ethiopia: Designing and Implementing a Rural Safety Net in a Low Income Setting: Lessons Learned from Ethiopia's Productive Safety Net Program 2005–2009.* Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/247601469672211732/Ethiopia-Designing-and-implementing-a-rur.

Ward, Patrick, Alex Hurrell, Aly Visram, Nils Riemenschneider, Luca Pellerano, Clare O'Brien, Ian MacAuslan, and Jack Willis. 2010. "Cash Transfer Programme for Orphans and Vulnerable Children (CT-OVC), Kenya." Oxford Policy Management, Oxford, UK.

Waze. 2018, 2020. "Driving Directions, Traffic Reports & Carpool Rideshares by Waze." 2018, 2020. https://www.waze.com/ccp.

Waze. 2018. "Waze Celebrates 600 Connected Citizens Program Partners." Blog, August 9. https://medium.com/waze/waze-celebrates-600-connected-citizensprogram-partners-36945fbceb66.

Wesolowski, Amy, Nathan Eagle, Andrew J. Tatem, David L. Smith, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2012. "Quantifying the Impact of Human Mobility on Malaria." *Science* 338 (6104): 267–70. https://doi.org/10.1126/science.1223467.

World Bank. 2015. *The Impact of Targeted Social Assistance on Labor Markets in Georgia: A Regression Discontinuity Approach.* Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/22502.

World Bank. 2019a. "Kosovo Social Assistance Scheme Study: Assessment and Reform Options." Report No. AUS0000628, World Bank, Washington, DC.

World Bank. 2019b. "A Quantitative Evaluation of the Greek Social Solidarity Income." World Bank, Washington, DC.

World Bank. 2020. "Strengthen Ethiopia's Adaptive Safety Net Project." PAD 3737, World Bank, Washington, DC.

World Bank. 2021a. "Social Assistance Programs in China: Current State and Directions for Reform." Technical Note P171959, Social Protection and Labor Markets Advisory Services and Analytics Programmatic ASA, World Bank, Washington, DC.

World Bank. 2021b. *World Development Report 2021: Data for Better Lives.* Washington, DC: World Bank. https://www.worldbank.org/en/results/2021/04/13/prioritizing-the-poorest-and-most-vulnerable-in-west-africa-togo-s-novissi-platform-for-social-protection-uses-machine-l.

World Bank and UNICEF (World Bank and United Nations Children's Fund). 2020. "Core Diagnostics of the Social Protection System in Armenia." World Bank, Washington, DC. https://www.unicef.org/armenia/media/9286/file/Core%20Diagnostic%20of%20the%20Social%20Protection%20System%20in%20Armenia.pdf.

Zeller, Manfred, Joseph Feulefack, and Andreas Neef. 2006. "How Accurate Is Participatory Wealth Ranking (PWR) in Targeting the Poor? A Case Study from Bangladesh." International Association of Agricultural Economists Annual Meeting, August 12–18, Queensland, Australia.

Zhang, Yuan, Qingjun Chen, and Sheng Zhong. 2016. "Privacy-Preserving Data Aggregation in Mobile Phone Sensing." *IEEE Transactions on Information Forensics and Security* 11 (5): 980–92. https://doi.org/10.1109/TIFS.2016.2515513.

Zou, James, and Londa Schiebinger. 2018. "AI Can Be Sexist and Racist—It's Time to Make It Fair." *Nature* 559: 324–26 (accessed September 15, 2019), https://www.nature.com/articles/d41586-018-05707-8.

7

Measuring the Performance of Targeting Methods

Phillippe Leite and Priyanka Kanth

Introduction

Too often, a narrow set of measures, limited data, and incautious inferences are used to drive policy discussions, resulting in misleading or incomplete conclusions on the performance of the targeting method used for a program. The literature on the topic is vast[1] and this chapter builds on it. A common error is focusing too much on simplistic errors of inclusion and exclusion, especially without considering program size. It is preferable to use measures that consider the full distribution and multiple dimensions of the program. Judgments about the findings must consider any limitations to what can be observed in the evaluation data being used, often from household surveys, with respect to definitions of the unit of observation, sample, measures of well-being, timing of the observations, and so forth. It is also important to understand the program context, rules, and implementation procedures to draw appropriately nuanced conclusions.

The objective of this chapter is to help policy makers, program administrators, and their advisers understand which measurements are most suited to answer which policy questions and to be able to critically read analyses and pick up on weaknesses in the choice of indicators, data, or conclusions drawn. The chapter starts with an illustration of analysis of a real program and some of the nuances involved in developing a good understanding of its strengths and weaknesses with respect to the method used. The chapter

then moves to pedagogy that leads from simple to more complex measures and from looking at only one or two aspects of a program to developing a more well-rounded picture. The main text should be readable by a focused layperson, and the mathematical definitions and formulae are provided in annex 7A.

Illustrative Case Study on How to Avoid Spurious Interpretations

An illustrative case study can help in understanding how misguided conclusions can often result from applying unsuitable measures or data to studying the performance of a method. The case study was created to demonstrate several common missteps in conducting assessments, including applying different thresholds for eligibility in the analysis than used by the program; using binary measurements; not using the entire population distribution to assess performance; not taking into account process evaluations highlighting implementation issues that could affect targeting outcomes; and the inherent limitations of household surveys, such as sampling and the date of the survey.

Brazil's Bolsa Escola, a conditional cash transfer program, was created in 2001, with intended nationwide coverage of 5.9 million families, covering 10.7 million children ages 7–14 years.[2] The program provided a monthly transfer of R$15 (US$6) per child for up to three children per family, conditional on 90 percent monthly school attendance to families with per capita income less than R$90 (US$36). In 2004, the Bolsa Escola program was merged with three other programs to create the Bolsa Familia program. The Benefício de Prestação Continuada (BPC) program, which was launched in 1996, was intended to provide monthly cash benefits to all the elderly ages 65 and older and the disabled, in amounts equivalent to the minimum wage. The transfers were unconditional and made independently of contributions to the social security system. Both programs used means testing as the method to determine eligibility, and their overall rules, guidelines, and financing were established by the federal government. However, the municipalities played a very important role in implementation, outreach, intake, registration, and onboarding.

The Brazilian National Household Sample Survey (PNAD)[3] of 2004 was the first representative household survey that collected information on participation in the Bolsa Escola and BPC programs. The survey discussed here was selected to illustrate the importance of understanding and applying similar eligibility criteria, using a full population distribution in the analysis, and understanding the importance of implementation issues prior to conducting performance assessments.

Method Performance Looking Solely at the Poverty Lines

Brazil commonly used two relative poverty lines, the extreme poverty line, which in the Brazilian context in 2004 was approximately a quarter of the minimum wage[4] (R$65, US$22.40), and the poverty line, which was half the minimum wage (R$130, US$44.80). Nevertheless, the income eligibility threshold for the Bolsa Escola program was households with per capita income less than R$90 (US$31.03), an amount that falls between the poverty and extreme poverty lines. The 2004 data suggest that 9 percent of households were participating in the Bolsa Escola program, while 12 percent of households (a total of 52 million households) were living in extreme poverty and 27 percent in poverty (see table 7.1). Solely using poverty lines as the poverty thresholds for measuring inclusion and exclusion errors, estimates show exclusion errors of 67 and 74 percent for extreme poverty and poverty, respectively, and inclusion errors of 59 and 25 percent, respectively.

With these numbers, it seems that the Bolsa Escola program's performance in reaching the poor was rather weak, leaving out large shares.

Table 7.1   Inclusion and Exclusion Errors: All Households

Extreme poor (Y ≤ R$65)	Selected for Bolsa Escola	Not selected for Bolsa Escola	Total
Yes	**2,024,630**	**4,100,312**	**6,124,942**
row %	33	67	100
column %	41	9	12
No	**2,967,738**	**43,104,112**	**46,071,850**
row %	6	94	100
column %	59	91	88
Total	**4,992,368**	**47,204,424**	**52,196,792**
row %	10	90	100
column %	100	100	100

Poor (Y ≤ R$120)	Selected for Bolsa Escola	Not selected for Bolsa Escola	Total
Yes	**3,755,866**	**10,430,311**	**14,186,177**
row %	26	74	100
column %	75	22	27
No	**1,236,502**	**36,774,113**	**38,010,615**
row %	3	97	100
column %	25	78	73
Total	**4,992,368**	**47,204,424**	**52,196,792**
row %	10	90	100
column %	100	100	100

Source: Based on Brazil National Household Survey (PNAD) 2004.
Note: R$65 = US$22.4; R$120 = US$41.38; Y = household per capita income.

However, the true eligibility thresholds include an income cutoff that is different from the poverty lines, as well as categorical targeting for households with children ages 7–14 years, which represents only 36 percent of the total number of Brazilian households.

Method Performance Restricting the Analysis to Households with Children and Using the Program Eligibility Threshold

Restricting the analysis to households with children ages 7–14 years and using the program's R$90 (US$31.03) eligibility threshold results in more program-relevant analysis, as shown in table 7.2. The caseload of households below the eligibility threshold drops from 9.1 million to 5.4 million, meaning that about 40 percent of the households living below the program's eligibility threshold had no children at the Bolsa Escola age, so they should not be classified as exclusion error. The exclusion error would be 51 percent and inclusion error 41 percent, which are much lower than the ones presented in table 7.1. The performance of the method still seems underwhelming; however, by restricting the analysis to the "intended to be treated" population as opposed to arbitrary cutoffs, the assessment is at least significantly more accurate.

Nevertheless, this is not a fully nuanced picture of the performance of the Bolsa Escola means test, as it does not account for the full distribution or other social programs with which it may interact.

Method Performance Using the Full Population Distribution

The Bolsa Escola program was not the only social program in Brazil; non-Bolsa Escola beneficiaries within the eligible income threshold could be

Table 7.2   Households with Children Ages 7–14 Years in Brazil

Extreme poor (Y ≤ R$90)	Selected for Bolsa Escola	Not selected for Bolsa Escola	Total
Yes	**2,660,263**	**2,760,101**	**5,420,364**
row %	49	51	100
column %	59	19	29
No	**1,869,161**	**11,488,914**	**13,358,075**
row %	14	86	100
column %	41	81	71
Total	**4,529,424**	**14,249,015**	**18,778,439**
row %	24	76	100
column %	100	100	100

Source: Based on Brazil National Household Survey (PNAD) 2004.
Note: R$90 = US$31.03; Y = household per capita income.

beneficiaries of other social programs. Figure 7.1 ranks households according to the per capita income distribution (ventiles). The solid dark blue line is the estimated share of Bolsa Escola beneficiaries at each income level; the yellow line is the estimated share of BPC[5] beneficiaries; and the gray line is the estimated share of other programs. The dark blue dashed line is the subpopulation of households with children ages 7–14 years who receive the Bolsa Escola. It shows that Bolsa Escola coverage of families with children at each income level is larger than or at par with the cumulative coverage of all social programs. Thus, looking at the complete household distribution shows that in terms of total coverage, the Bolsa Escola program outdoes all the other programs. Further, the beneficiary incidence analysis for Bolsa Escola shown in figure 7.2 indicates that most of the budget is spent on households living on less than the minimum wage per capita, and about 25 percent of the beneficiaries are above the poverty line.

That being said, the program still has errors that are large enough to be worth trying to correct. Why is it that the program was not able to target more precisely? Often the answer lies in the implementation of different delivery chain components.

Figure 7.1   Social Program Coverage in Brazil
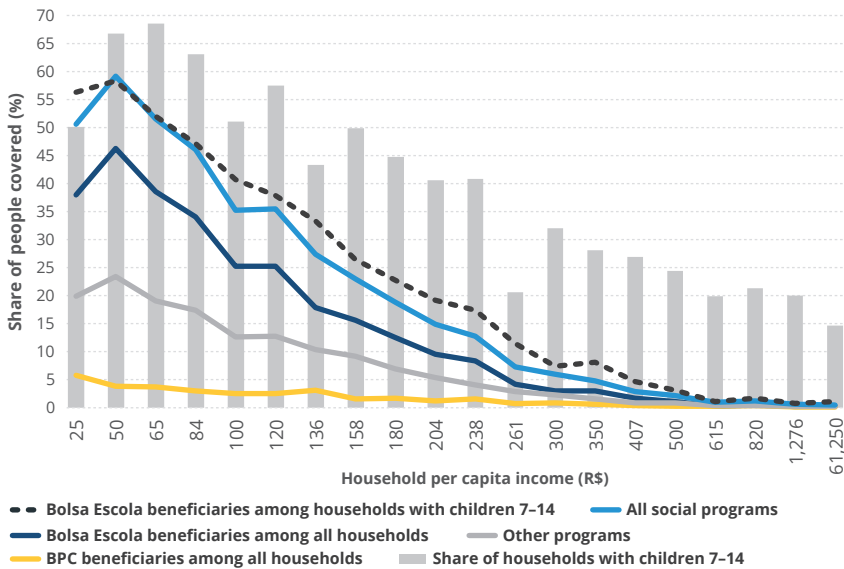
Source: Based on Brazil National Household Survey (PNAD) 2004.
Note: BPC = Benefício de Prestacão Continuada.

Figure 7.2   Bolsa Escola Beneficiary Incidence, 2004
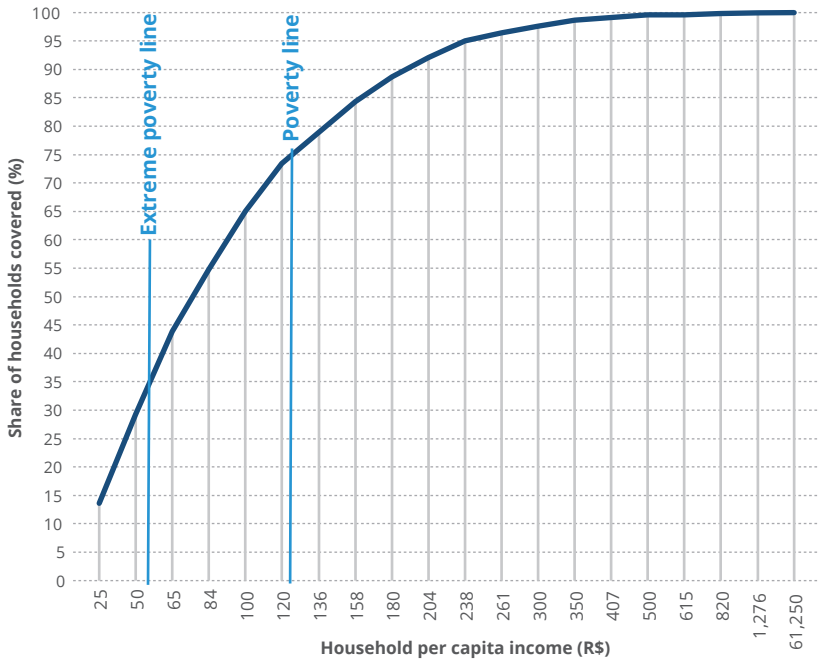
Source: Based on Brazil National Household Survey (PNAD) 2004.

Role of Implementation in Ensuring Efficiency

Implementation of the Bolsa Escola program was accompanied by studies on program assessment to establish a body of evidence on the effectiveness of the program. De Janvry et al. (2005) and Pinheiro do Nascimiento and Aguiar (2006)[6] provide thorough discussions of implementation challenges that affected the effectiveness of the program, including the efficiency of the method in including the desired population. The first key thing to note is that the although the Bolsa Escola program was a federal program, it was implemented completely by the municipalities. The decentralized nature of the implementation resulted in variations in implementation across municipalities, due to (1) variations in identification of beneficiaries, (2) variations in beneficiary selection, (3) variations in monitoring and enforcement of conditionalities, and (4) social controls over program implementation. Both studies show variation in the implementation capacity of Brazilian municipalities given their different sizes, remoteness, connectivity, and so on. These implementation variations have substantial impacts on program

implementation, which are often reflected in poor outcomes. This is not necessarily due to the design of the program or the method used; it could be due to how the delivery chain components were implemented.

Summary of Elaborations or Corrections of the Initial Limited Analysis of the Outcomes of the Bolsa Escola Case Study

- Using the right income eligibility cutoff as opposed to arbitrarily using the poverty line
- Restricting the sample to households within the income eligibility to those that also comply with the categorical requirements of the method, namely having children in a certain age group
- Checking for the coexistence of other mutually exclusive social programs that might explain low take-up of a specific program
- Understanding coverage over the entire population distribution
- Understanding the implementation challenges, since it is not always possible to measure them
- Using a data set that allows for population-wide analysis of program coverage and incidence.

Some Aspects That Could Affect the Method's Performance but Often Cannot Be Measured, Especially in Poorer Countries or, in the Case of Brazil, Poorer Municipalities

- The year of the survey may be more recent than the year in which the most recent survey sweep/large-scale recertification of beneficiaries was done; thus, some households that were eligible when assessed have prospered in a long-run sense or had a good year.
- Families may be excluded due to failures in the delivery system—limited outreach, high transaction costs, lack of identification (ID), and so forth.
- An eligibility assessment improperly classifies some households.

What to Look for When Conducting Method Assessments

1. When measuring errors of inclusion and exclusion, it is vital to use the same threshold as used for eligibility rather than a more generous definition of poverty.

As evidenced in the illustrative case study on the Bolsa Escola program in Brazil, using different eligibility cutoffs for analysis than those used in the program design can lead to inaccurate assessments of accuracy and coverage of the eligible population. To illustrate this, consider an economy of 10 individuals, a poverty line of $20, and a government desire to

reduce poverty. In this economy (table 7.3), the poverty headcount is estimated at 40 percent, and the government has allocated $30 for a transfer program. The observed total income gap is $40, which means that the current budget allocated would not suffice to eradicate poverty. As the budget is smaller than the needs, policy administrators decided that it would be better to use this budget to reach the poorest 20 percent (the relatively poorest for sure, perhaps labeled the extreme poor) instead of all the poor. They would like to compare such a program with a universal program that might avoid social tensions and promote solidarity across the population. Four scenarios are presented as follows:

- Scenario 1 depicts a process that has perfect inclusion of the poorest two individuals.
- Scenario 2 depicts a process with all benefits going to poor individuals but to only one of the two extreme poor individuals.
- Scenario 3 depicts a process with all benefits going to poor individuals but missing the two poorest individuals.
- Scenario 4 reaches all individuals.

Table 7.3 illustrates the economy and presents the two most commonly used metrics for measuring performance: exclusion and inclusion errors (see the formulae in annex 7A). First, the errors are estimated against two benchmarks: the real poverty line and a relative poverty line that was used as the eligibility threshold (the poorest 20 percent).

Table 7.3   Inclusion and Exclusion Errors in a 10-Person Economy

Person	1	2	3	4	5	6	7	8	9	10	Total	
Observed income ($)	6	8	11	15	21	25	30	40	50	100	306	
Income gap ($)	14	12	9	5	0	0	0	0	0	0	40	
Poor	1	1	1	1	0	0	0	0	0	0	4	
Poorest 20%	1	1	0	0	0	0	0	0	0	0	2	
Scenario 1	1	1	0	0	0	0	0	0	0	0	2	**PL: EE: 50%; IE: 0%** P20: EE: 0%; IE: 0%
Scenario 2	0	1	1	0	0	0	0	0	0	0	2	**PL: EE: 50%; IE: 0%** P20: EE: 50%; IE: 50%
Scenario 3	0	0	1	1	0	0	0	0	0	0	2	**PL: EE: 50%; IE: 0%** P20: EE: 100%; IE: 100%
Scenario 4	1	1	1	1	1	1	1	1	1	1	10	**PL: EE: 0%; IE: 60%** P20: EE: 0%; IE: 80%

Source: Original compilation for this publication.

Note: EE = exclusion error; IE = inclusion error; PL = poverty line; P20 = relative poverty line (the poorest 20 percent).

Although there is a rationale for each of the benchmarks, the numbers and their interpretations are different. Using the benchmark of the 40 percent poverty line finds very high exclusion rates that may be read as poor performance of the method when the problem is that the size of the program is smaller than the size of the population of interest. A program for 20 percent of the population when 40 percent of the population is below the poverty line will have at least 50 percent exclusion errors, as it can only cover one-half of the poor population even if all those included are poor. Conversely, it is easier for small programs to avoid inclusion error. In this economy, measured against the poverty line, the three small program scenarios have no or smaller inclusion errors despite the 50 percent exclusion error as program size for the first three scenarios is smaller than the poor population. At the same time, looking only for exclusion error hides the significant inclusion error in scenario 4, which is 60 percent.

Mistakenly looking at the poverty line as benchmarking, the first three scenarios have similar errors. Looking at the actual targeted population, that is, the bottom 20 percent, reveals nuances in performance. Scenario 1 has a perfect outcome with 0 percent errors as designed, and scenario 3 has the worst outcome with 100 percent error. Therefore, the example illustrates that exclusion and inclusion errors cannot be analyzed in absence of program size, and results vary depending on the benchmark set.

2. Even in the binary world of measuring method performance, there are measures preferable to simple errors of inclusion and exclusion.

Combining three indicators—coverage and inclusion, and exclusion errors—helps to normalize the errors accounting for the program size and measure performance according to Ravallion (2007) and Wiesmann et al. (2009). Ravallion (2007) shows that the *targeting differential (TD)* is a metric that is easy to interpret and brings the coverage of the intended population into the equation. When only the intended group gets help from the social program and all of them are covered, the TD equals 1, which is the measure's upper bound; when only the unintended group gets the program and all of them do, the TD equals −1, the lower bound. Hence, the TD ranges between −1 (low performance) and 1 (high performance) (table 7.4). Detailed formulae are provided in annex 7A.

As the TD is estimated as the difference between the coverage of the eligible population (100 − exclusion error) and the inclusion error, it can be estimated for the four scenarios and for both the real poverty line and the poorest 20 percent benchmark. Scenario 1, which reaches the poorest two individuals, would have a TD of 50 percent for the full poverty line and 100 percent for the poorest 20 percent, implying perfect inclusion for the

Table 7.4   Targeting Differential

		PL			P20		
	Program size	EE	IE	TD	EE	IE	TD
Scenario 1 (%)	20	50	0	50	0	0	100
Scenario 2 (%)	20	50	0	50	50	50	0
Scenario 3 (%)	20	50	0	50	100	100	−100
Scenario 4 (%)	100	0	60	40	0	80	20

Source: Original compilation for this publication.
Note: EE = exclusion error; IE = inclusion error; PL = poverty line; P20 = relative poverty line (the poorest 20 percent); TD = targeting differential.

real poverty line. Scenario 3, which misses the two poorest individuals, has a TD of −100 percent for the poorest 20 percent. For the largest program, scenario 4, the zero exclusion error does not imply perfect method performance according to the TD. For the real poverty line, scenario 4 would have the worst performance among the four cases, and for the poorest 20 percent, its performance is close to that of scenario 2, which has exclusion errors and is far from the perfect case in scenario 1.

Wiesmann et al. (2009) suggest looking at the errors in a different way to assess performance as the indicators are clearly related since a larger number of individuals correctly classified as eligible means coverage improvements and reduction of inclusion errors, when holding fixed the size of the population to be protected by the social program. This case is illustrated by the improvements that occur when moving from scenario 3 to scenario 2 to scenario 1. Coverage increases from 0 to 50 and then to 100 percent as inclusion error drops from 100 to 50 to 0 percent for the poorest 20 percent. Increasing the number of good matches at the expense of large increases in the number of targeted units does not help, as scenario 4 shows. In scenario 4, the increase in targeted units from 2 to 10 succeeds in reaching all of the poorest 20 percent, but inclusion errors reach 80 percent.

Wiesmann et al. (2009) present other indicators generated from the same 2 x 2 table, *specificity* and *positive predictive value*. The specificity suggests good performance by looking at the rate of proper bad matches, that is, the number of noneligible households that are properly classified as noneligible. The positive predictive value measures the good matches, that is, the eligible who are correctly selected as eligible. A well-performing program will have not only high coverage of the intended population, but also high specificity and high positive predictive value when suitable eligibility criteria are chosen to determine participation. The detailed formulae are presented in annex 7A.

Table 7.5   **Coverage, Specificity, and Positive Predictive Value for Benchmarking against the Poverty Line**

Scenarios 1–3	Selected as participant	Not selected	Total
Poor	2	2	4
Nonpoor	0	6	6
Total	2	8	10

Coverage equals 2/4 = 0.5 (50%)
Specificity equals 6/6 = 1 (100%)
Positive predicted value equals 2/2 = 1 (100%)

Scenario 4	Selected as participant	Not selected	Total
Poor	4	0	4
Nonpoor	6	0	6
Total	10	0	10

Coverage equals 4/4 = 1 (100%)
Specificity equals 0/6 = 0 (0%)
Positive predicted value equals 4/10 = 0.4 (40%)

Source: Original compilation for this publication.
Note: MCC = Matthews correlation coefficient.

These two indicators—specificity and positive predictive value—are estimated for the scenarios in tables 7.5 and 7.6. For scenarios 1 to 3 and 4, the estimations confirm that scenarios 1 to 3 have better performance than scenario 4 regardless of the benchmarking. Especially considering the poorest 20 percent benchmark, scenarios 1 and 3 are the best and worst cases, respectively, as before, but scenario 2 shows better performance indicators than scenario 4.

In the machine learning literature, other metrics (also detailed in annex 7A), such as the *Matthews correlation coefficient (MCC)*, F_1 score, and F_2 score, are largely used to measure the quality of binary classification. The MCC was introduced by biochemist Brian W. Matthews[7] and can also be estimated for measuring precision. Hence, it can be used here for measuring performance. It measures the correlation coefficient between the observed and predicted binary classifications. The higher is the correlation between the true and predicted values, the better is the prediction. When there are only good matches, the value of the MCC is 1, indicating perfect positive correlation. Conversely, when the classifier always misclassifies and there are only bad matches, the value of the MCC is −1, representing perfect negative correlation. The MCC value is always between −1 and 1, with 0 meaning that the classifier is no better than a random flip of a fair coin. Estimating the MCC on the poorest 20 percent benchmarking for scenarios 1 to 3 confirms that they have good performance (table 7.7). The MCC cannot be estimated for scenario 4. Scenarios 1 and 3 remain the best and worst cases, respectively.

Table 7.6   Coverage, Specificity, and Positive Predictive Value for Benchmarking against the Poorest 20 Percent

Scenario 1	Selected as participant	Not selected	Total
Poor	2	0	2
Nonpoor	0	8	8
Total	2	8	10

Coverage equals 2/2 = 1 (100%)
Specificity equals 8/8 = 1 (100%)
Positive predicted value equals 2/2 = 1 (100%)

Scenario 2	Selected as participant	Not selected	Total
Poor	1	1	2
Nonpoor	1	7	8
Total	2	8	10

Coverage equals 1/2 = 0.5 (50%)
Specificity equals 7/8 = 0.875 (87.5%)
Positive predicted value equals 1/2 = 0.5 (50%)

Scenario 3	Selected as participant	Not selected	Total
Poor	0	2	2
Nonpoor	2	6	8
Total	2	8	10

Coverage equals 0/2 = 0 (0%)
Specificity equals 6/8 = 0.75 (75%)
Positive predicted value equals 0/2 = 0 (0%)

Scenario 4	Selected as participant	Not selected	Total
Poor	2	0	2
Nonpoor	8	0	8
Total	10	0	10

Coverage equals 2/2 = 1 (100%)
Specificity equals 0/8 = 0 (0%)
Positive predicted value equals 2/10 = 0.2 (20%)

Source: Original compilation for this publication.

The F_a score also measures accuracy. It is estimated on errors of inclusion and errors of exclusion. It is a harmonic mean of measures of the positive predicted values and the bad matches. The F_a score reaches its best value at 1 and worst at 0. The values of a that are commonly used are a = 1 and a = 2. F_1 is intuitively a balanced measure of inclusion and exclusion errors; F_2 is a measure of inclusion and exclusion errors with double weight on the latter and on the positive predicted values.

Estimating F_a for a = 1 and a = 2 on the poorest 20 percent benchmarking for scenarios 1 to 3 and 4 confirms that scenarios 1 to 3 have good performance. The MCC cannot be estimated for scenario 4. Scenarios 1 and 3 remain the best and worst cases, respectively, as for both MCC and F_1. However, as there is no exclusion error for Scenario 4,

Table 7.7 MCC, F_1, and F_2 Benchmarking against Poorest 20%

Scenario 1	Selected as participant	Not selected	Total
Poor	**2**	**0**	2
Nonpoor	**0**	**8**	8
Total	2	8	10

MCC equals {(**2** × **8**) − (**0** × **0**)}/{square root of (**2** + **0**) × (**2** + **0**) × (**0** + **8**) × (**0** + **8**)} = 1 (100%)
F_1 score equals (1 + 1) × **2** / {((1 + 1) × **2**) + 1 × **0** + **0**} = 1 (100%)
F_2 score equals (1 + 4) × **2** / {((1 + 4) × **2**) + 4 × **0** + **0**} = 1 (100%)

Scenario 2	Selected as participant	Not selected	Total
Poor	**1**	**1**	2
Nonpoor	**1**	**7**	8
Total	2	8	10

MCC equals {(**1** × **7**) − (**1** × **1**)} / {square root of (**1** + **1**) × (**1** + **1**) × (**1** + **7**) × (**1** + **7**)} = 0.375 (37.5%)
F_1 score equals (1 + 1) × **1** / {((1 + 1) × **1**) + 1 × **1** + **1**} = 0.5 (50%)
F_2 score equals (1 + 4) × **1** / {((1 + 4) × **1**) + 4 × **1** + **1**} = 0.5 (50%)

Scenario 3	Selected as participant	Not selected	Total
Poor	**0**	**2**	2
Nonpoor	**2**	**6**	8
Total	2	8	10

MCC equals {(**0** × **6**) − (**2** × **2**)} / {square root of (**0** + **2**) × (**0** + **2**) × (**2** + **6**) × (**2** + **6**)} = −.25 (−25%)
F_1 score equals (1 + 1) × **0** / {((1 + 1) × **0**) + 1 × **2** + **2**} = 0 (0%)
F_2 score equals (1 + 4) × **0** / {((1 + 4) × **0**) + 4 × **2** + **2**} = 0 (0%)

Scenario 4	Selected as participant	Not selected	Total
Poor	**2**	**0**	2
Nonpoor	**8**	**0**	8
Total	10	0	10

MCC equals {(**2** × **0**) − (**0** × **8**)} / {square root of (**2** + **8**) × (**2** + **0**) × (**8** + **0**) × (**0** + **0**)} = n.a.
F_1 score equals (1 + 1) × **2** / {((1 + 1) × **2**) + 1 × **0** + **8**} = 0.33 (33%)
F_2 score equals (1 + 4) × **2** / {((1 + 4) × **2**) + 4 × **0** + **8**} = 0.56 (56%)

Source: Original compilation for this publication.
Note: MCC = Matthews correlation coefficient.

increasing the weights on both the positive predictive values and exclusion errors leads to higher precision for F_2 compared with F_1 as the exclusion error is null.

Essama-Nssah (2018) proposes other measures, such as the targeting success rate and the agreement coefficient built from the same 2 x 2 table. The targeting success rate measures the rate of cases properly classified (eligible poor and noneligible nonpoor) in the population, that is, the sum of the main diagonal divided by the population. The agreement coefficient accounts for the fact that the social programs' true and false cases can

happen by chance. The author proposes testing the chance that participation in the program occurs by chance instead of the method. The agreement coefficient varies from −1 to 1, and the value 1 indicates perfect agreement while 0 is the expected value when agreement is purely by chance. The Landis and Koch (1977) table helps to interpret the agreement statistic as follows: a coefficient less than 0.20 means a poor targeting outcome; between 0.21 and 0.40, a fair targeting outcome; between 0.41 and 0.60, a moderate targeting outcome; and greater than 0.61, good performance, as presented in annex 7A.

In the 10-person economy example earlier in this section, the estimation for the benchmark poverty line showed moderate performance for scenarios 1 to 3 and poor performance for scenario 4 (table 7.8). For the poorest 20 percent, scenario 1 performed well and scenario 3 performed poorly, as expected, but for scenario 2, the agreement coefficient is estimated at 0.375, meaning a fair performance.

Lindert, Skoufias, and Shapiro (2006) bring in the value of the transfer while still basing performance measurement on only inclusion and exclusion errors. The rationale behind their case is that a program can be the following:

• Effective in absolute terms because benefits reach a significant share of the desired population, but ineffective in relative terms since the

Table 7.8   Benchmarking the Agreement Coefficient against the Poverty Line

Scenarios 1–3	Selected as participant	Not selected	Total
Poor	2	2	4
Nonpoor	0	6	6
Total	2	8	10

Targeting success rate (TSR) equals (2 + 6) / 10 = 0.8 (80%)
Joint probability equals {(2 × 4) / (10 × 10) + (8 × 6) / (10 × 10)} = {0.8 + 0.48} = **0.56**
The agreement coefficient equals {(0.8 − **0.56**) / (1 − **0.56**)} = 0.54

Scenario 4	Selected as participant	Not selected	Total
Poor	4	0	4
Nonpoor	6	0	6
Total	10	0	10

Targeting success rate (TSR) equals (4 + 0) / 10 = 0.4 (40%)
Joint probability equals {(10 × 4) / (10 × 10) + (0 × 6) / (10 × 10)} = {0.4 + 0} = **0.4**
The agreement coefficient equals {(0.4 − **0.4**)/(1 − **0.4**)} = 0

Source: Original compilation for this publication.

benefits are too small to improve the household welfare of the desired population.
- Ineffective in absolute terms because the benefits reach a significant share of the nondesired population, but effective in relative terms as the benefits suffice to improve the household welfare of the desired population.

Table 7.9 illustrates this case for the poorest 20 percent in the 10-person economy, focusing only on scenarios 2 and 4 benchmarked at the poorest 20 percent. As the budget that was allocated was $30, scenario 2 implies a transfer of $15 per person, while scenario 4 implies a transfer of $3 per person. Scenario 4 is effective in absolute terms as all poor people are beneficiaries (coverage of the desired population is 100 percent). However, the $3 transfer is insufficient to bring a single individual above the poverty line, as presented in the row *income after transfer*. In contrast, scenario 2 has 50 percent coverage of the desired population as the benefit reaches a nondesired case, individual 3. So, scenario 2 has certain inefficiency in absolute terms, but it is effective in relative terms as the selected poor recipient (individual 2) got a benefit that allowed crossing the poverty line (the formula is provided in annex 7A).

3. It is preferable to consider a program's performance over the whole welfare distribution.

A significant weakness/bias when assessing performance with only errors of inclusion/exclusion or indices derived from those is not accounting for how close or far above or below people are from a given benchmark, as was

Table 7.9 Relative Efficiency of Programs

Person	1	2	3	4	5	6	7	8	9	10	Total
Observed income ($)	6	8	11	15	21	25	30	40	50	100	306
Income gap ($)	14	12	9	5	0	0	0	0	0	0	40
Transfer scenario 2 ($)	0	15	15	0	0	0	0	0	0	0	30
Income after transfer ($)	6	**23**	**26**	15	21	25	30	40	50	100	336
Transfer scenario 4 ($)	3	3	3	3	3	3	3	3	3	3	30
Income after transfer ($)	**9**	**11**	**14**	**18**	24	28	33	43	53	103	336

Source: Original compilation for this publication.

evident in the illustrative case study on the Bolsa Escola program in Brazil. The exclusion of an extremely poor person can/should be more important than the exclusion of a person whose income is just below the poverty line. The inclusion of someone who is above but close to the poverty line can/should be more acceptable than the inclusion of someone among the far richer population. However, the metrics discussed in the prior sections flag only whether a person's income is above or below the eligibility line, not how far above or below, which is insufficient from a redistribution point of view. Presenting results on coverage, benefit (beneficiary) incidence, and generosity across the full distribution is therefore a richer way to consider the issue. Figure 7.3 illustrates results from a proxy means-tested program designed to cover 17 percent of the population in a given at a time when the poverty rate was 30 percent. The figure shows that 65 percent of the poorest decile receives the program, as well as 39 percent of the population in the second decile and that 60 percent of the benefits accrue to the poorest 61 percent population (38 plus 23). Though there are errors of inclusion and exclusion, both coverage and the share of benefits accruing to each decile drop significantly beyond the fourth decile. Errors of inclusion are concentrated in the lower part of the distribution where they are most acceptable; errors of exclusion are higher closer to the eligibility threshold where they are less alarming than at the bottom of the distribution.

To deepen the analysis of redistribution, Coady and Skoufias (2004) recommend applying the *Distribution Characteristic Index (DCI)*—which was

Figure 7.3   Illustration of Coverage and Benefit Incidence for a Program for the Poorest 15 Percent of the Population
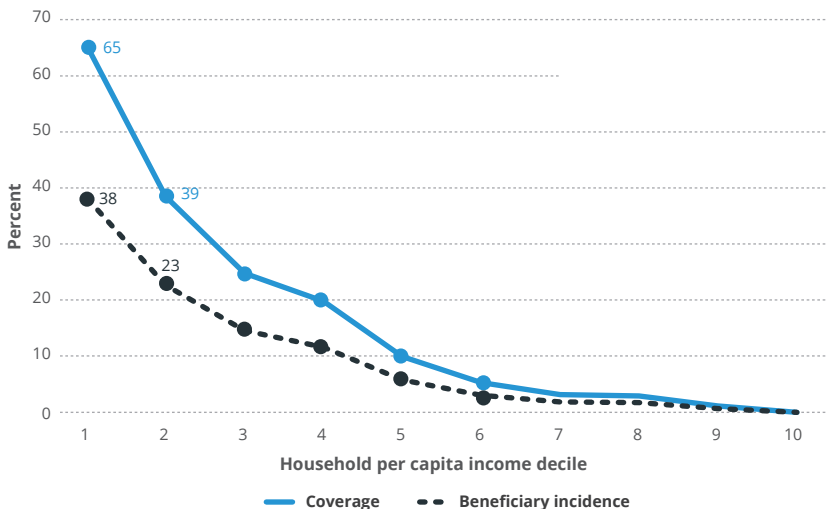
Source: Original compilation for this publication.

developed for taxation analysis by Ahmad and Stern (1991) and Newbery and Stern (1987)—to social program transfers. The DCI makes value judgments when comparing programs because it allows for the quantitative comparison of how much better or worse programs are relative to each other independently of the (different) sizes of their budgets.[8] The DCI also avoids the controversy and difficulty of specifying a poverty line and accounts for the fact that a nonpoor household participating in the program may be just above the poverty line and not in the top of the welfare distribution. The DCI measures the change in social welfare generated for each dollar of transfer budget distributed. A positive (negative) redistributive efficiency value implies that a progressive (regressive) adjustment needs to be made to allow for differentiation of transfers across households. Moreover, the DCI can be decomposed into two other indicators—efficiency and redistribution—and it is estimated for different measures of the sensitivity to inequality (epsilon) factor, with values often presented for epsilon ranging from 0.5 (low sensitivity) to 2 (high sensitivity).

Table 7.10 illustrates the DCI, using the same scenarios as in table 7.9 and a fifth scenario, a perfectly targeted program to eradicate poverty (scenario 5) with transfers set at the income gap (a sort of guaranteed minimum income). As the DCI measures the change in social welfare (marginal benefit) achieved by transferring a standardized budget (say, $1), the perfectly targeted, gap-filling scenario 5 would be more effective as a redistribution policy (targeting efficiency) and when the monthly value of transfers that poor people receive exceeds the monthly value of transfers that wealthy people receive (redistributive efficiency). Decomposing the

Table 7.10   Distributional Characteristic Index and Its Decomposition

Epsilon	0.5	1.0	1.5	2.0
DCI				
Scenario 5	1.56	2.49	4.06	6.73
Scenario 2	1.46	2.16	3.20	4.78
Scenario 4	1.04	1.25	1.69	2.49
Targeting efficiency				
Scenario 5	1.48	2.25	3.51	5.61
Scenario 2	1.46	2.16	3.20	4.78
Scenario 4	1.04	1.25	1.69	2.49
Redistributive efficiency				
Scenario 5	0.08	0.25	0.55	1.12
Scenario 2	0	0	0	0
Scenario 4	0	0	0	0

Source: Original compilation for this publication.
Note: DCI = Distributional characteristic index.

indicator into targeting and redistribution confirms better efficiency as it has higher redistributive impact even for scenario 2 against scenario 4. The universal program under scenario 4 still generates acceptable efficiency, but it is lower than the other two scenarios.

A limitation of the DCI is comparability across countries and time; however, the measure can be normalized under some circumstances. The DCI measure is the sum of the utility distributed to beneficiaries, weighted by their pretransfer income relative to a benchmark. This has consequences for comparability between countries at any particular time or for the same country in different years. If a program has the same beneficiaries in two consecutive years but the shape of the welfare distribution changes, the DCI will change (due to a change in the social welfare weights), which does not reflect any change in the method's performance. Comparability can be enhanced through normalization; the DCI under the perfect targeting scenario can be calculated—households are sorted from poorest to richest and benefits are distributed on this basis—and the DCI of actual method performance can be presented as a percentage of perfect. A limitation of this approach is that the DCI rewards any households receiving benefits; thus, bad performance still leads to a positive measure even if it is less valued. A comprehensive normalization thus requires establishing not only a perfect targeting DCI, but also a perfect mistargeting DCI in which households are ranked from richest to poorest and benefits are disbursed on this basis. The program's actual performance is thus the extent to which it moves from perfect mistargeting toward perfect targeting.

However, depending on how a targeted program is implemented, normalization may not be possible. How a program is implemented will affect the ability to normalize. A program can (1) determine all beneficiaries as those with a means test, hybrid means test, or proxy means test score below a certain threshold, regardless of how many beneficiaries this is (the "line" approach; see chapter 8); or (2) determine beneficiaries as those who have the lowest means test, hybrid means test, or proxy means test scores up until a program quota is met (the "quota" approach). Under a quota approach, perfect mistargeting can be estimated: it is the DCI when the richest X = quota households are selected. Under the line approach, the model may identify more or less than the true number of poor as eligible. If the number is less, the program DCI can never be 100 percent, even if the model ranks the households in correct order of welfare. If the model identifies more households as poor than the true number, the program DCI can be greater than the perfect DCI because there are more beneficiaries than considered under the perfect scenario.

A program's impacts on poverty and redistribution are a function of three factors: coverage, the way in which resources are distributed (incidence/progressivity), and the relative size of the transfers compared with

the well-being of the poor (generosity). This "triangle" allows measuring whether a program succeeds in reaching the intended population and provides a large enough benefit to have an impact on their lives. To reduce the poverty headcount, benefits must be larger than the income gap to the poverty line. The depth of poverty and distribution can be improved by smaller transfers, but smaller transfers have smaller impacts. Therefore, two indicators (see the detailed formulae in annex 7A) are needed to assess redistribution[9]:

- The *benefit (beneficiary) incidence indicator* is estimated as the proportion of transfers (beneficiaries) received in each group.
- The *generosity indicator* is estimated as the value of the transfers received by a group divided by the total welfare aggregate of beneficiaries in that group.

The benefit (beneficiary) incidence indicator has the advantage of being independent of the size of the program. In other words, it allows measuring whether among the beneficiaries the intended population is the more prevalent, while generosity indicators measure how significant the size of the transfer is for the welfare of the group, regardless of the program's size.

The performance triangle measures the performance of a social program as a function of the program's accuracy in reaching the intended population, coverage of the intended population, and the importance of transfers relative to the level of welfare without the program transfers. As such, it can be used to answer the following questions. What share of the beneficiaries is indeed correctly identified? What proportion of the intended population group is covered or served by the program? Would the program have impacts on poverty and inequality?

Here, the performance triangle is illustrated using the 10-person economy. For benchmarking at the poorest 20 percent, coverage of the intended group is estimated at 100 percent for the new scenario 5, 50 percent for scenario 2, and 100 percent for scenario 4. The benefit incidence indicators for the three scenarios are 65 percent (35 plus 30) for scenario 5, 50 percent (0 plus 50) for scenario 2, and 20 percent (10 plus 10) for scenario 4. Generosity for each case is 1.86 (2.3 and 1.5 times for each individual) for scenario 5, 1.07 (0 and 1.88 times for each individual) for scenario 2, and 0.43 (0.50 and 0.47 times for each individual) for scenario 4 times the group's current income. Plotting the three indicators in the performance triangle, the shape of the perfect case (scenario 5) is better mimicked by scenario 2 than by scenario 4, and scenario 2 is also expected to have more redistribution than scenario 4 (table 7.11 and figure 7.4). Therefore, this example illustrates that reaching the intended population but not providing an adequate benefit is insufficient for redistribution.

Table 7.11   Benefit Incidence and Generosity, the Poverty Triangle

Person	1	2	3	4	5	6	7	8	9	10	Total
Observed income ($)	6	8	11	15	21	25	30	40	50	100	306
Transfer scenario 5 ($)	14	12	9	5	0	0	0	0	0	0	40
Share of benefits (%)	35	30	22	12	0	0	0	0	0	0	100
Generosity	2.33	1.5	0.82	0.45	0	0	0	0	0	0	0.13
	(26/14) = 1.86										
Transfer scenario 2 ($)	0	15	15	0	0	0	0	0	0	0	30
Share of benefits (%)	0	50	50	0	0	0	0	0	0	0	100
Generosity	0	1.88	1.36	0	0	0	0	0	0	0	0.1
	(15/14) = 1.07										
Transfer scenario 4 ($)	3	3	3	3	3	3	3	3	3	3	30
Share of benefits (%)	10	10	10	10	10	10	10	10	10	10	1
Generosity	0.50	0.37	0.27	0.20	0.14	0.12	0.10	0.07	0.06	0.03	0.10
	(6/14) = 0.43										

Source: Original compilation for this publication.

Figure 7.4   Performance Triangle
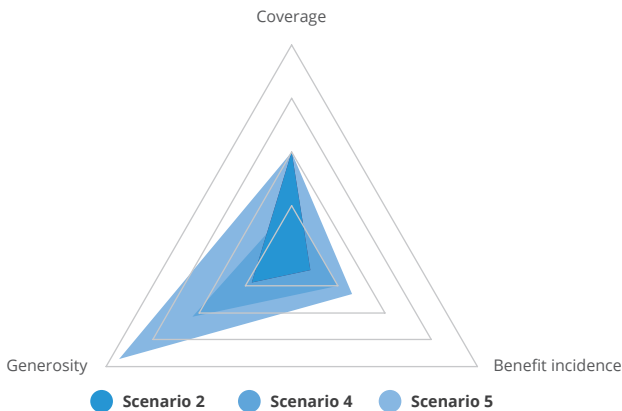
Source: Original compilation for this publication.

4. (Simulated) impacts on poverty and inequality are important metrics.

Since the justification for targeting is often to increase the impact on poverty and/or inequality for a given budget, it is pertinent to use measures of poverty and inequality to throw light on the choice of methods. Poverty measures can be used to test the relevance of these methods for measuring a program's impacts on poverty and inequality. Two of the most commonly used poverty measures are the headcount index, FGT(0), and the poverty gap index, FGT(1),[10] and the Gini index is used to measure inequality. The FGT(0) is the most popular measure among policy makers, but it is insensitive to income gains or losses unless the individual crosses the poverty line as a result of a transfer. Thus, it fails to capture the benefit of lowering the poverty gap among those who remain poor, which is a problem solved by using FGT(1).

To determine the effect on poverty or inequality requires dealing with the issue of the counterfactual. To establish the impact of a social protection program on poverty, indicators without the program, $FGT(0,1)_A$ and $Gini_A$, should be compared with indicators with the program, $FGT(0,1)_B$ and $Gini_B$. The problem is that often only one of these two states can be observed. If the program already exists, then $FGT(0,1)_B$ and $Gini_B$ can be calculated from the household survey data, but $FGT(0,1)_A$ and $Gini_A$ cannot be calculated because data are only available for the situation with the program, and there is no information before an individual became a beneficiary of the program. For simulating a program, the problem is reversed: $FGT(0,1)_A$ and $Gini_A$ can be seen in the data but not the actual $FGT(0,1)_B$ and $Gini_B$, only a simulation assuming that the program is the only new thing that affects the indicators.

A common approach is to subtract all transfers received under the program to get a different income measure for each person, $y_i - t_i$, where t_i is the amount of transfer that person i received. Recalculating the poverty measure with this value yields $FGT(0,1)_A$ and $Gini_A$. Comparing $FGT(0,1)_B$ and $Gini_B$ with $FGT(0,1)_A$ and $Gini_A$ gives an estimate of the program's poverty and inequality impacts. The findings rest on the assumption that receiving the transfer does not change people's behavior in a way that would affect income. If people were to work less in response to the transfer payment, then $(y_i - t_i)$ would not be the correct value for income in the absence of the program; their income would be somewhat higher than that. In an extreme case, behavioral responses might completely offset the effect of the transfer on income, so that y_i is the correct estimate of pre-transfer income, and the program should be seen as having no effect on poverty. The literature on impact evaluations that use control groups to solve the counterfactual problem with fewer assumptions shows that

recipients of social assistance transfers generally maintain their work effort and thus presumably their income; indeed, sometimes they increase these (see the discussion in chapter 2).

The challenge is more significant for countries that use consumption or expenditure to measure welfare. For an existing program, consumption or expenditure cannot be observed exclusive of any transfers that the program provides without an assumption that the individuals consume x percent of the transfer. Therefore, all transfers received under the program cannot be subtracted to get a different welfare measure for each person, c_i–t_i, where t_i is the amount of transfer that person i received. In this case, it is important to assume the value x or estimate the income/consumption elasticity to estimate c_i–xt_i as a counterfactual. Unfortunately, there is often no good way to estimate each person's counterfactual from a single cross-sectional data set.[11] In some instances, the assumption that there is no behavioral response to the social protection program may be a good one, while in others, it may be far from the truth.

Acknowledging the limitations, some sense of the impacts of the program on poverty and inequality can still be established. A perfectly targeted program may have no impact on poverty rates if the benefit or coverage is too small, but it can have an impact on the poverty gap and inequality as it brings people closer to the poverty line. From a targeting point of view, reductions of the FGT(1) and Gini are more relevant. The benefit-cost ratio is a measure that shows how much of a $1 transfer goes toward reducing the poverty gap. The benefit-cost ratio ranges from 0 to 1, and 1 is the upper bound where all transfers go to poverty gap reduction.

Using the 10-person economy and three scenarios in table 7.12, it can be shown that scenario 5, which transfers the precise income gap to each poor individual, fully reduces both the poverty headcount and the poverty gap and has a cost-benefit ratio of 1, while the Gini coefficient would drop 29 percent. Scenario 2 moves two of the four poor individuals from poverty, leading to a 50 percent reduction in the poverty headcount. The impact on the poverty gap is 53 percent as the two individual beneficiaries are not among the poorest. The total amount of money transferred per individual

Table 7.12 Impacts on Poverty and Inequality

	ΔFGT(0) (%)	ΔFGT(1) (%)	ΔGini (%)	Cost-benefit ratio
Scenario 5	100	100	29	1
Scenario 2	50	53	18	0.70
Scenario 4	0	30	9	0.40

Source: Original compilation for this publication.
Note: FGT(0) = poverty headcount; FGT(1) = poverty gap.

would move them above the poverty line, but both of them would receive more money than needed to reach the line. As such, the cost-benefit ratio is 0.70 ($21 was needed and $30 was transferred), and the Gini drops by 18 percent. For scenario 4, there is no impact on the poverty headcount as the transfers were too small, but there is some reduction in the poverty gap (30 percent) as $12 reached the poor of the $30 transferred, representing a cost-benefit ratio of 0.4, and the Gini dropped by 9 percent.

To conclude, many assessments are presented as functions of inclusion and exclusion errors, ignoring the welfare distribution or impacts on poverty and inequality. Exclusion and inclusion errors are rather blunt measures and miss much of the redistributive impacts of social transfers, which are important features of social programs. Both errors are also calculated with different thresholds for different countries/programs, which makes benchmarking difficult. Furthermore, they often do not take fully into account the specificities of the program's objectives, which may include goals other than poverty reduction. Finally, most such analyses or simulations are based on national household surveys, making little use of data from impact evaluations or process evaluations that may shed light on elements that are important for the success or failure of design or implementation.

5. Comparison of performance across programs of different design must consider multiple performance indicators.

This subsection provides an illustration of a performance assessment using a real household budget survey. Tables 7.13 to 7.16 summarize some of the results for two programs observed in the data, benchmarking against the extreme poverty line. The current poverty rate, FGT(0), in the exercise is 10 percent, FGT(1) is 2.3 percent, and the Gini inequality measure is 0.283.

The initial assessment compares the exclusion and inclusion errors of programs A and B. If the main concern is errors of exclusion, then program A seems more acceptable. If it is errors of inclusion, program B would be preferred. These indicators do not convey the full picture of the programs and their potential impacts, so it is worth digging a bit deeper.

Table 7.13   Exclusion and Inclusion Errors

	Exclusion error (%)	Inclusion error (%)
Program A	55.7	88.7
Program B	63.7	49.9

Source: Original compilation for this publication.

Table 7.14   Expanding Binary Indicators

	Exclusion error (%)	Inclusion error (%)	Coverage of the extreme poor (%)	Targeting differential	Share of total population covered (%)
Program A	55.7	88.7	44.3	−0.44	20.4
Program B	63.7	49.9	36.4	−0.13	3.7

Source: Original compilation for this publication.

Table 7.15   Impact on Welfare

	Benefit incidence (%)	Generosity	Coverage of the poor (%)	DCI
Program A	12.5	0.12	44.3	0.65
Program B	56.3	0.29	36.4	1.99

Source: Original compilation for this publication.
Note: DCI = distributional characteristic index.

Table 7.16   Impacts on Poverty and Inequality

	ΔFGT(0) (%)	ΔFGT(1) (%)	ΔGini (%)	Cost-benefit ratio
Program A	5	9	1	0.469
Program B	2	17	1	0.875

Source: Original compilation for this publication.
Note: FGT(0) = poverty headcount; FGT(1) = poverty gap.

The introduction of coverage of the poor and program size sheds some useful light on the analysis. Program A has larger coverage of the extreme poor population, 44 percent, but at the same time, it is much larger than program B. This result corroborates Cornia and Stewart's (1993) point that it is easier to ensure that no benefits go to the nonpoor in a small program than in a large one, so low inclusion errors can be biased in favor of small programs. The TD measure by the subtraction of inclusion error from the coverage of the extreme poor population indicates that despite its size, program B seems to be better targeted as the TD is smaller, but it is still far from the upper bound of perfect outcomes.

The addition of indicators to measure the impact on redistribution shows again a different picture of performance. Programs A and B have different triangle shapes, with lower generosity and less progressive incidence for program A, despite higher coverage rates (figure 7.5). The results indicate that the smaller program size and higher exclusion error can be offset by

Figure 7.5   Performance Triangle for Two Programs
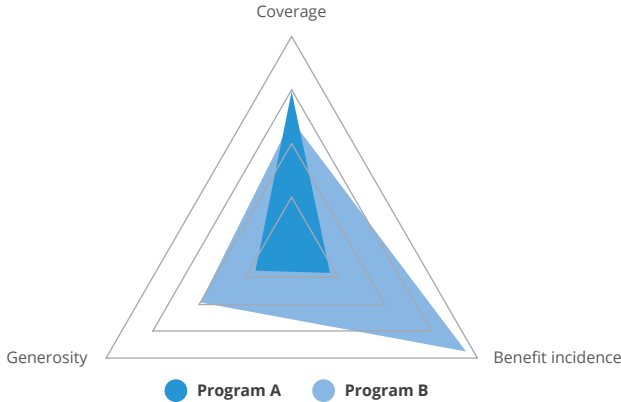
Source: Original compilation for this publication.

the impacts on welfare caused by the higher generosity and much more progressive incidence in program B.

A closer look at the redistributive impacts through the DCI confirms that despite being small, program B has more progressive incidence and higher generosity than program A, and program B seems to have a higher transfer for those in the bottom tail of the welfare distribution compared with those in the upper tail. The fact that program A' DCI is less than 1 implies that the change in social welfare (marginal benefit) achieved by transferring a standardized budget (say, $1) through the program is not efficient for program A, as less than $1 would go to changing the welfare of the poor.

Despite the challenges of having a proper counterfactual in the absence of the programs, table 7.16 assumes that all transfers were consumed by the household and each counterfactual is estimated independently of the other program. Program B's higher generosity among the poor leads to a higher impact on FGT(0) combined with the higher program coverage of the poor. However, as shown by the DCI, program B has more redistributive power and more progressivity, which translate into a higher impact on FGT(1). This impact is corroborated by the almost double cost-benefit ratio of program B compared with that of program A.

In conclusion, by looking only at exclusion errors, program A—a universal child allowance—would have better targeting performance than program B—a guaranteed minimum income program. When other indicators are considered, the small, narrowly targeted guaranteed minimum income would show greater redistributive impact and targeting performance than the child allowance and a larger impact on the poverty gap, FGT(1). The smaller impact on the poverty headcount, FGT(0), occurs because the

guaranteed minimum income reaches those far from the poverty line. The higher coverage of the child allowance manages to reach those close to the poverty line who easily cross the poverty line with a small transfer. The larger impact of the child allowance on the reduction in the poverty head-count comes at the cost of covering 5.5 times more people and spending twice as much as the guaranteed minimum income program. The findings show the poverty- and distribution-related considerations for a policy maker. However, policy can have other objectives as well. In the country from which this example is drawn, the child allowance also has an objective of increasing the birth rate among women of childbearing age, and the analysis explained here is inadequate to shed light on that objective or the weighting between the poverty and fertility impacts of the programs. Further, the analysis is not conclusive about whether a guaranteed mini-mum income or a child allowance program is better. That depends on the weights given to different factors, and indeed many countries have both programs as they serve different policy niches.

6. The most common basis for measuring targeting outcomes is household surveys, but they suffer from some limitations.

The explosion of the availability of household survey data in the past 20 or 30 years has been a huge boon to understanding the performance of social programs. In many countries, consumption and expenditure surveys, Living Standards Measurement Study surveys, multiple indicator cluster surveys, Demographic and Health Surveys, or similar have become a part of the statistical infrastructure with some regularity and credibility and are now a part of the expected toolkit of policy analysis. The random sampling frames and representativeness of household surveys are important to be able to make statements about distribution.

However, the design of questionnaires and often inadequate or poorly representative sample sizes or the type or absence of welfare collected by household surveys may limit their ability to cast full light on the perfor-mance of targeted social assistance programs,[12] due to nonsampling and sampling errors.[13] The questionnaires of many such surveys contain ques-tions that are relevant for only a subset of social programs, and the questions are not always adapted to the specificities of different programs. Household surveys can suffer from poor question wording, definitional differences between the nature of the indicator and the way the question is asked, misunderstandings on the parts of both the interviewer and the inter-viewed, lack of knowledge on the program received (for example, the respondent is not aware of the benefit amount or frequency of a particular program received by another household member), inability of the inter-viewed to keep up with program name changes,[14] and deliberate

misreporting of welfare or program participation, which can also be associated with the respondents' misunderstanding of the objectives of household surveys and the fear that these may be government audits/spot checks of program beneficiaries.[15] Moreover, very often, information is not disaggregated by individual programs, the transfer amount is not collected, or public and private transfers are mixed in the same questions or collected for a different assistant unit.[16] Survey estimations can also be imprecise due to sampling errors if the population of interest is not adequately represented in the sampling frame, which causes loss in the statistical precision of any indicator derived from the household survey. The problem may be especially marked in countries where social assistance programming has been small or consisted of start/stop programs so that statistical institutes would not have been able or expected to capture social assistance programming in their surveys.[17] Moreover, sample sizes or designs may be a problem especially in capturing information about small social protection programs.[18] Therefore, it is important to estimate the standard errors[19] of each indicator produced when comparing the performance of countries or programs to show differences through calculating confidence intervals, to determine the precision of the indicators and compare different programs. Household surveys may not measure the same concept of welfare as programs use—focusing on income rather than consumption or vice versa or focusing on the same concept but with different degrees of thoroughness on issues such as own-produced food or treatment of owner-occupied housing.[20] Household Income and Expenditure Surveys, Living Standards Measurement Study surveys, and the like will have detailed income or consumption modules on which to base distributive analysis, but Demographic and Health Surveys have only limited asset information on which to build an index of wealth.

Finally, household data are available only periodically; thus, they capture the welfare of households at a different time than that when eligibility assessments for any programs they benefit from would have been done. Therefore, data from household surveys may cast some light on some aspects of the overall success of the larger and more stable programs, but they are not the same as tests of actual eligibility decision-making processes. For those, audits, simulations, process evaluations, and/or decision-contemporary special-purpose surveys are needed. For example, suppose there is a direct cash transfer program for those living below the poverty line. The determination of eligibility for the program is done through a means test, and the program rules are clear that the selected beneficiary will receive 36 months of transfers. Suppose that a recipient household uses the transfers to make small investments (for example, in poultry) and after a year such investments start bringing extra income to the household. Two years later, when the household survey data collection

is done, the household's income may be above the eligibility threshold because the household is still receiving transfers regularly according to the program rules and earning extra income from the small business (poultry). Should this case be tagged as inclusion error? Suppose that another household that has not been poor over the past years experiences an income shock a few days before the national survey data collection and the family has yet to apply for a social program to receive support. In the household survey data, the family will appear as poor and uncovered, an exclusion error, although it has not been erroneously deemed ineligible and it might even be too soon for there to be real concern that the household was excluded due to inadequate outreach or stigma and transaction costs. These two cases highlight the different issues between using static data from a period that is different from when the eligibility assessment was done to make conclusions about the accuracy of eligibility assessment.

Finally, household data are available only periodically and thus capture the welfare of households at a time different from that when eligibility assessments for any programs they benefit from would have been done.

Concluding Remarks

Several factors must be considered in conducting a proper assessment of the methods used to determine individuals' or households' eligibility for social assistance. It is important to relate data to the program design as closely as possible—using the eligibility criteria (income threshold, family composition, location, and so forth). Ideally, the assessment should use a range of measurements and consider the distribution over the whole population. Often it is important to triangulate among data from administrative records, random sample household surveys, process evaluations, or impact evaluations and to be aware of the limitations of any of these.

Annex 7A: Formulae for the Indicators

Basic Measurements

| | | Predicted | | |
		Participants	**Nonparticipants**	**Total**
Actual	Eligible	True positive - n_{11}	False negative n_{12}	$n_1 = n_{11} + n_{12}$
	Noneligible	False positive - n_{21}	True negative n_{22}	$n_2 = n_{21} + n_{22}$
	Total	$n_1 = n_{11} + n_{21}$	$n_2 = n_{12} + n_{22}$	$n = n_{11} + n_{12} + n_{21} = n_{22}$

Note: n_1 = total eligible population; n_2= total noneligible population; n_1 = total participants; n_2 = total nonparticipants; n_{11} = true positive; n_{12} = false negative; n_{21} = false positive; n_{22} = true negative.

Inclusion error is the false positives divided by the participant population: $\dfrac{n_{21}}{(n_{11} + n_{21})}$.

Exclusion error is the false negatives divided by the eligible population: $\dfrac{n_{12}}{(n_{11} + n_{12})}$.

Coverage is estimated as (1 – exclusion error) or the true positives divided by the eligible population: $\dfrac{n_{11}}{(n_{11} + n_{12})}$.

Targeting differential = $\left(\dfrac{n_{11}}{(n_{11} + n_{12})} \right) - \left(\dfrac{n_{21}}{n_1} \right)$.

Specificity is the true negatives divided by the noneligible population: $\dfrac{n_{22}}{n_2}$.

Positive predicted value is the true positives divided by the participant population: $\dfrac{n_{11}}{n_{.1}}$.

Misclassification is the sum of the secondary diagonal divided by the total population: $\dfrac{(n_{12} + n_{21})}{n}$.

Targeting success rate is the sum of the main diagonal divided by the total population: $\dfrac{(n_{11} + n_{22})}{n}$.

The Matthews correlation coefficient is a correlation coefficient between the observed and predicted binary classifications:

$$\frac{(n_{11} + n_{22}) - (n_{12} + n_{21})}{\sqrt{n_{11} + n_{21})\,(n_{11} + n_{12})(n_{21} + n_{22})(n_{12} + n_{22})}} \; .$$

The F_a score is measured as $F_a = \dfrac{\left(1 + a^2\right)n_{11}}{\left(1 + a^2\right)n_{11} + \left(a^2\right)n_{12} + n_{21}}$.

Essama-Nssah (2018) Agreement Coefficient

The agreement coefficient accounts for the fact in social programs, the true and false cases can happen by chance. The test measures whether a large share of true and false cases did not happen by chance. The kappa coefficient is a chance-corrected agreement coefficient that removes the amount due to chance, that is, *if participation is independent of eligibility, their joint probability is equal to the product of the marginal probabilities*, meaning that p_e equals

$$p_e = \left[\left(\frac{n_{c1}}{n}\right) \times \left(\frac{n_{r1}}{n}\right) + \left(\frac{n_{c2}}{n}\right) \times \left(\frac{n_{r2}}{n}\right)\right] \text{where}$$

$$n_{cj} = \left(n_{1j} + n_{2j}\right) \; for \; j = 1, 2$$

$$n_{ri} = \left(n_{i1} + n_{i2}\right) \; for \; i = 1, 2$$

and the chance-agreement correction entails subtracting this value from p_a, which equals the targeting success rate (see above). Furthermore, $(1 - p_e)$ indicates the amount of agreement that is not expected to occur by chance. Hence, the kappa coefficient is defined as follows:

$$k = \frac{p_a - p_e}{1 - p_e} \; .$$

The k statistics vary from −1 to 1. A value of 1 indicates perfect agreement, while 0 is the expected value when agreement is purely by chance. A negative value indicates systematic disagreement (Viera and Garrett 2005). The strength of agreement indicated by the kappa statistic can be determined on the basis of the Landis-Koch scale.

Landis and Koch (1977) Interpretation of the Kappa Statistic

0–0.20	Poor
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Incidence Analysis

	1	2	3	4	Total
Benefits allocated	b_{11}	b_{12}	b_{13}	b_{14}	b_1
Benefit incidence	$b_1 = \dfrac{b_{11}}{b_{1.}}$	$b_2 = \dfrac{b_{12}}{b_{1.}}$	$b_3 = \dfrac{b_{13}}{b_{1.}}$	$b_4 = \dfrac{b_{14}}{b_{1.}}$	1
Participants	n_{11}	n_{12}	n_{13}	n_{14}	n_1
Beneficiary incidence	$p_1 = \dfrac{n_{11}}{n_1}$	$p_2 = \dfrac{n_{12}}{n_1}$	$p_3 = \dfrac{n_{13}}{n_1}$	$p_4 = \dfrac{n_{14}}{n_1}$	1
Total welfare	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_1
Total welfare of participants	Y_{11}^p	Y_{12}^p	Y_{13}^p	Y_{14}^p	$Y_{1.}^p$
Generosity	$g_1 = \dfrac{b_{11}}{Y_{11}^p}$	$g_2 = \dfrac{b_{12}}{Y_{12}^p}$	$g_3 = \dfrac{b_{13}}{Y_{13}^p}$	$g_4 = \dfrac{b_{14}}{Y_{14}^p}$	$g_1 = \dfrac{b_1}{Y_1^p}$

Note:
b_1 = total benefits allocated to the program;
b_{1i} = total benefits allocated to group *l*;
n_1 = total eligible population of the program;
n_{1i} = total eligible population belonging to group *i*;
Y_1 = total welfare of the population;
Y_{1i} = total welfare of group *i*;
$Y_{1.}^p$ = total welfare of program participants;
Y_{1i}^p = total welfare of program participants in group *i*.

Distributional Characteristics Index

The social welfare impact of any transfer program is then:

$$dW = \sum_h \frac{\partial W}{\partial V^h} \frac{\partial V^h}{\partial m^h} dm^h \equiv \sum_h \beta^h dm^h$$

where β^h (the welfare weight) is the social value of extra income to household h; $V^h (p,y)$ is the indirect utility function for household h; p is the vector of commodity prices faced by the household; and y is total household income. A transfer program can be characterized by a vector $dm = [dm^1,\ldots, dm^h,\ldots dm^H]$, where $dm^h > 0$ for beneficiary households and $dm^h = 0$ for nonbeneficiary households.

Multiplying and dividing the right-hand side of the equation by the program budget $B = \Sigma_h dm^h$ gives

$$dW = \sum_h \beta^h \frac{dm^h}{\sum_h dm^h} \sum_h dm^h \equiv \sum_h \beta^h \theta^h \sum_h dm^h \equiv \lambda B$$

where θ^h is the share of the transfer budget going to each household and $\lambda = \sum_h \beta^h \theta^h$.

The term λ, also called the distributional characteristic index (or DCI) of the program, represents the marginal benefit of distributing a unit of income ($1) through a transfer program relative to the marginal cost (the budget).

The distributional characteristic is a weighted average of the welfare weights of the social welfare impact of a transfer instrument multiplied by the share of the transfer going to each household. Therefore, it differs across transfer programs because the welfare weights differ across households and the structure of the transfers (that is, who receives them and how much) differs across programs. The greater is the proportion of the budget ending up in the hands of the poorest households, the greater is the distributional characteristic.

Thus, the calculation of λ requires specifying the welfare weights for each household. A useful and common method for specifying these weights derives from Atkinson's (1970) constant elasticity social welfare function. In that function, the relative welfare weight of household h is calculated as:

$$\beta^h = \left(Y^k \big/ Y^h \right)^\varepsilon$$

where k is a reference household. Often that reference household is on the poverty line z, so $y^k = z$.

In this equation, ε captures aversion to inequality, with aversion increasing in ε. For example, $\varepsilon = 0$ implies no aversion to inequality—a dollar has a dollar of value regardless of who receives it—so all the welfare weights take on the value of unity. A value $\varepsilon = 1$ implies that if household h has twice (half) the income of household k, then the welfare weight of household h is 0.5 (2.0), but the welfare weight of household k is unity. As ε approaches infinity, the welfare impact of transfers to the poorest households dominates the evaluation, consistent with a Rawlsian maxi-min social welfare perspective where one cares only about the welfare impact on the poorest households.

The distributional characteristic can be decomposed into two indices; each index is conceptually and empirically useful. Define dm^* as the average transfer to beneficiaries, that is, the total amount of transfers divided by the number of beneficiaries, where beneficiaries are those with $dm^h > 0$. Then add and subtract dm^* across all beneficiaries, so for all nonbeneficiaries, $dm^* = 0$, to get

$$\lambda = \frac{\sum_h \beta^h dm^*}{\sum_h dm^h} + \frac{\sum_h \beta^h \left(dm^h - dm^* \right)}{\sum_h dm^h} = \lambda_T + \lambda_R$$

where λ_T is the *targeting efficiency* and λ_R is the *redistributive "sizing" efficiency* of the transfer instrument.

The variable λ_T captures the welfare impact of a program that divides B into equal amounts and gives them to the same beneficiary households, and λ_R is the adjustment that needs to be made to allow for the differentiation of transfer sizing across households in a more progressive ($\lambda_R > 0$) or regressive ($\lambda_R < 0$) manner.

For programs that give every beneficiary identical transfers, the uniform transfers $\lambda_R = 0$. The sense in which λ_R captures the redistributive efficiency of the policy instrument is made clearer by interpreting it as the welfare impact of a self-financing program that transfers dm^h to households and finances the transfers by a lump sum tax on all beneficiary households, that is, all households with $dm^h > 0$.

Poverty

This chapter uses the Foster, Greer, and Thorbecke (1984) family of poverty headcount ($\alpha = 0$), poverty gap ($\alpha = 1$), and poverty severity ($\alpha = 2$).

$$FGT(\alpha) = \frac{\sum_{h=1}^{n}\left(1 - \dfrac{Y^{h}}{Z}\right)^{\alpha} w^{h}}{\sum_{h=1}^{n} w^{h}}$$

if $y^{h} \leq Z$,

where y^{h} is household h's well-being, Z is the poverty line, and w^{h} is household h's expansion factor.

To simulate the $FGT(\alpha)$ without transfers, y^{h} is replaced by $y^{h} - t^{h}$, where t^{h} is the benefit amount received by household h.

Gini

The Gini coefficient is given by

$$Gini = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left| Y^{i}w^{i} - Y^{j}w^{j} \right|}{2\sum_{i=1}^{n}\sum_{j=1}^{n} y^{j}w^{j}}$$

where y^{h} is household h's well-being and w^{h} is household h's expansion factor.

To simulate the Gini without transfers, y^{h} is replaced by $y^{h} - t^{h}$, where t^{h} is the benefit amount received by household h.

Cost-Benefit Ratio

The cost-benefit ratio is estimated as the division of the total difference between the poverty gap with and without benefits, and the total benefit amount. It is represented mathematically as follows:

$$A = \sum_{h=1}^{n}\left(1 - \frac{Y^{h}}{Z}\right)w^{h} \ if \ y^{h} \leq Z$$

$$B = \sum_{h=1}^{n}\left(1 - \frac{Y^{h} - t^{h}}{Z}\right)w^{h} \ if \ y^{h} - t^{h} \leq Z$$

$$C = \sum_{h=1}^{n} t^{h} w^{h}$$

$$Cost\text{-}benefit \ ratio = \frac{B - A}{C}.$$

Normalized Root Mean Square Error

In statistical modeling and particularly regression analyses, a common way to measure the quality of the fit of the model is the root mean square error (RMSE)[21] (also called root mean square deviation). It is estimated as the square root of the ratio of the quadratic sum of the differences between the observations and their predicted values, divided by the number of observations.

When the predicted responses are very close to the true responses, the RMSE will be small. If the predicted and true responses differ substantially—at least for some observations—the RMSE will be large.

A value of zero would indicate a perfect fit to the data.

In machine learning approaches, the model performance must be compared using training and test data sets (after the modification). In this case, we calculate the normalized RMSE (NRMSE).

Normalizing the RMSE facilitates the comparison of data sets or models with different scales. There are various methods of RMSE normalizations in the literature.

RMSE can be normalized by:

- The mean: NRMSE equals the ratio of RMSE and the average value of the observed dependent variable, *y*.
- The difference between maximum and minimum: NRMSE equals the ratio of RMSE and the difference between the maximum and minimum values of *y*.
- The standard deviation: NRMSE equals the ratio of RMSE and the standard deviation of *y*.
- The interquartile range: NRMSE equals the ratio of RMSE and the difference between the 25th and 75th percentiles of the distribution of *y*.

Notes

1. The literature reviewed in this chapter includes Coady, Grosh, and Hoddinott (2004); Jones, Vargas, and Villar (2008); Kidd (2011); Garcia and Moore (2012); Paes-Sousa, Regalia, and Stampini (2013); Apella and Blanco (2015); Brown, Ravallion, and van de Walle (2016); Kidd, Gelders, and Bailey-Athias (2016); Devereux et al. (2017); Hanna and Olken (2018); and Kidd and Athias (2020).
2. See Pinheiro do Nascimiento and Aguiar (2006).
3. PNAD, which was carried out annually, was finished in 2016, with the release of information for 2015. Planned to produce results for Brazil, Major Regions, Federation Units and nine Metropolitan Regions (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba, and Porto Alegre), it surveyed, on an ongoing basis, general characteristics of the

population, education, labor, income, and housing, and, according to the information needs for the country, having the household as its unit of survey. The PNAD was replaced, with updated methodology, by the Continuous National Household Sample Survey, or Continuous PNAD, which provides more comprehensive territorial coverage and quarterly short-term information on the workforce nationwide. During the 49 years of its existence, PNAD has been an important instrument for the formulation, validation, and evaluation of policies targeted at the socioeconomic development of the population and the improvement of living conditions in the country (https://www.ibge.gov.br/en /statistics/social/population.html).

4. The minimum wage in 2004 was set at R$260 (US$89.60).
5. The BPC program was launched in 1996 to provide monthly cash benefits to all elderly people ages 65 and older and the disabled, in amounts equivalent to the minimum wage. Transfers were unconditional and made independently of contributions to the social security system.
6. https://unesdoc.unesco.org/ark:/48223/pf0000150153.
7. Matthews (1975).
8. Annex 7A provides the detailed formula for the DCI.
9. The detailed formulae are provided in annex 7A.
10. See Foster, Greer, and Thorbecke (1984).
11. The data needed to estimate such a counterfactual income are quite demanding. See Ravallion (2003) for an intuitive introduction.
12. See Leite, Rodriguez Alas, and Reboul (2019).
13. Sampling and nonsampling errors determine the degree to which a survey statistic differs from its true value due to imperfections in the way the statistic is collected, and both affect the measurement of performance. Sampling error (coverage errors) exist when an inadequate sampling frame leads to indicator under- or over-coverage, leading to underrepresentation of social program coverage in household surveys. Nonsampling errors exist due to poor question wording, definitional differences between the nature of the indicator and the way the question is asked, misunderstandings from both the interviewer and the interviewed, and deliberate misreporting that can also be associated with the miscomprehension of household survey and government audits/spot checks over program beneficiaries. In addition, when programs are too small, the sampling frame often misses them and the performance indicators will be biased. Developing and regularly updating a comprehensive set of comparable and accessible indicators using administrative data, household surveys, and impact evaluation surveys are needed for analysis of program performance.
14. In many countries, program names change, sometimes without proper communication. For example, the conditional cash transfer program in Mexico was originally called PROGRESA, then changed to Oportunidades, and finally became PROSPERA. In Colombia, the System for the Selection of Beneficiaries for Social Programs (SISBEN) was created as a program for health subsidies but then became the social registry for social programs; however, many elderly

beneficiaries still refer to receiving SISBEN instead of receiving the Regimen Subsidiado de Salud. Questionnaires can also be misleading due to poor wording about addressing the program's new name.

15. To minimize such errors, a well-designed survey facilitates a high proportion of valid responses. Questions and response options should be worded in common terms used by respondents, rather than using legal terminology or technical statistical terms. In addition, interviewers should be properly trained and understand the nature of the social programs. They should also be able to identify the right respondents for the different questions. Any potential issue should be anticipated and included in the interview manual.

16. When the assistance unit is smaller than the household, it may be difficult to make proper inferences about the program if social program data are collected only for the household. For example, individual programs' coverage and incidence may not be estimated if program participation and benefits are only collected at the household level.

17. Therefore, a good understanding of programs and sampling can inform the researcher about the limitation and reliability of the indicators generated. In general, due to the costs of bad planning, national statistical offices may be able to increase the sample size in particular zones so that household survey data are representative of the particular area or program to improve estimation precision. If such an approach for minimizing such sampling errors is not viable, programs may have to be aggregated with other programs of the same nature to increase their statistical robustness, although program specificity would be lost. Further, when generating estimations, it is important to analyze the size of the standard errors produced for each indicator.

18. In many countries, household surveys have some issues of representativeness in the poorest areas or rural areas, meaning that the ability to evaluate some programs, small or large ones, and to measure their performance can be affected. Checking the quality of survey information against administrative sources is good practice but does not solve the problem if surveys are not redesigned. Small programs—that is, those that cover a small proportion of the total population—are hard to capture by means of nationally representative surveys. The estimated coverage of such programs will be imprecise, because the sample size of a typical household survey is not large enough to capture enough beneficiaries.

19. The estimation of standard errors must be done using the sampling weights after setting the correct sampling design for estimation.

20. Ravallion (2008) discusses the issue of the impact of nonsampling errors on mistargeting of the Chinese Dibao program by showing that the problem is that the concept of poverty underlying a program's objectives often appears to be broader than the way income is normally defined and measured from surveys. Hence, the program's apparent mistargeting could simply reflect the fact that the survey-based measure of income is not a sufficient statistic for deciding who is poor.

21. See Otto (2019).

References

Ahmad, Etisham, and Nicholas Stern. 1991. "The Theory and Practice of Tax Reform in Developing Countries." Cambridge University Press. https://EconPapers.repec.org/RePEc:cup:cbooks:9780521265638.

Apella, Ignacio, and Gastón Blanco. 2015. "Effectiveness of Targeting Mechanisms Utilized in Social Protection Programs in Bolivia." Social Protection and Labor Discussion Paper 1506, World Bank, Washington, DC.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2 (3): 244–63. ISSN 0022-0531. https://doi.org/10.1016/0022-0531(70)90039-6.

Brown, Caitlin, Martin Ravallion, and Dominique van de Walle. 2016. "A Poor Means Test? Econometric Targeting in Africa." NBER Working Paper No. 22919, National Bureau of Economic Research, Cambridge, MA. JEL No. I32, I38, O15.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/14902.

Coady, David, and Emmanuel Skoufias. 2004. "On the Targeting and Redistributive Efficiencies of Alternative Transfer Instruments." *Review of Income and Wealth* 50: 11–27. https://doi.org/10.1111/j.0034-6586.2004.00109.x.

Cornia, Giovanni Andrea, and Frances Stewart. 1993. "Two Errors of Targeting." Innocenti Occasional Papers EPS 36, United Nations Children's Fund, Florence, Italy.

De Janvry, Alain, Frederico Finan, Elisabeth Sadoulet, Donald Nelson, Kathy Lindert, Benedicte de la Brière, and Peter Lanjouw. 2005. "Brazil's Bolsa Escola Program: The Role of Local Governance in Decentralized Implementation." Social Protection Discussion Paper 0542, World Bank, Washington, DC.

Devereux, Stephen, Edoardo Masset, Rachel Sabates-Wheeler, Michael Samson, Althea-Maria Rivas, and Dolf te Lintelo. 2017. "The Targeting Effectiveness of Social Transfers." *Journal of Development Effectiveness* 9 (2): 162–211. http://dx.doi.org/10.1080/19439342.2017.1305981.

Essama-Nssah, Boniface. 2018. "Assessing the Performance of Targeting Mechanisms." ECINEQ Working Paper 2018-457, Society for the Study of Economic Inequality, https://ideas.repec.org/p/inq/inqwps/ecineq2018-457.html.

Foster, James, Joel Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–65.

Garcia, Marito, and Charity M. T. Moore. 2012. *The Cash Dividend: The Rise of Cash Transfer Programs in Sub-Saharan Africa*. Washington, DC: World Bank.

Hanna, Rema, and Benjamin A. Olken. 2018. "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries." Working Paper 24939, National Bureau of Economic Research, Cambridge, MA.

Jones, Nicola, Rosana Vargas, and Eliana Villar. 2008. "Cash Transfers to Tackle Childhood Poverty and Vulnerability: An Analysis of Peru's Juntos Programme." *Environment and Urbanization* 20 (1): 255–73.

Kidd, Stephen. 2011. "Targeting the Poorest: An Assessment of the Proxy Means Test Methodology." Australian AID, Canberra, Australia. https://doi.org/10.13140/2.1.3567.6488.

Kidd, Stephen, and Diolá Athias. 2020. "Hit and Miss: An Assessment of Targeting Effectiveness in Social Protection with Additional Analysis." Working Paper, Development Pathways, Sidcup, UK.

Kidd, Stephen, Bjorn Gelders, and Diolá Bailey-Athias. 2016. "Exclusion by Design: The Effectiveness of the Proxy Means Test." ESS Working Paper No. 56, International Labour Office, Development Pathways, Sidcup, UK.

Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–74.

Leite, Phillippe, Claudia P. Rodriguez Alas, and Virginie Reboul. 2019. "Measuring Social Protection and Labor Programs through Household Surveys." World Bank, Washington, DC.

Lindert, Kathy, Emmanuel Skoufias, and Joseph Shapiro. 2006. "Redistributing Income to the Poor and the Rich: Public Transfers in Latin America and the Caribbean." Social Protection Discussion Paper 0605, World Bank, Washington, DC.

Matthews, Brian W. 1975. "Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme." *Biochimica et Biophysica Acta* 405 (2): 442–51.

Newbery, David M. G., and Nicholas Stern (eds.). 1987. *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.

Otto, Saskia A. 2019. "How to Normalize the RMSE." Blog post. https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/.

Paes-Sousa, Romulo, Ferdinando Regalia, and Marco Stampini. 2013. "Conditions for Success in Implementing CCT Programs: Lessons for Asia from Latin America and the Caribbean." Social Protection and Health Division Policy Brief No. IDB-PB-192, Inter-American Development Bank, Washington, DC.

Pinheiro do Nascimiento, Elmer, and Marcelo Aguiar. 2006. "Bolsa Escola: historia y evolución." UNESCO. Instituto Internacional de Planeamiento de la Educación, IIEP Document code: iiep/web doc/2006.05. https://unesdoc.unesco.org/ark:/48223/pf0000150153.

Ravallion, Martin. 2003. "The Debate on Globalization, Poverty and Inequality: Why Measurement Matters." *International Affairs* 79 (4): 739–54.

Ravallion, Martin. 2007. "How Relevant Is Targeting to the Success of an Antipoverty Program?" Policy Research Working Paper 4385, World Bank, Washington, DC.

Ravallion, Martin. 2008. "Miss-Targeting or Miss-Measurement?" *Economics Letters* 100: 9–12.

Viera, Anthony J., and Joanne M. Garrett. 2005. "Understanding Interobserver Agreement: The Kappa Statistic." *Family Medicine* 37: 360–63.

Wiesmann, Doris, Lucy Bassett, Todd Benson, and John Hoddinott. 2009. "Validation of the World Food Programme's Food Consumption Score and Alternative Indicators of Household Food Security." Discussion Paper 00870, Poverty, Health, and Nutrition Division, International Food Policy Research Institute, Washington, DC.

8

Machine Learning and Prediction of Beneficiary Eligibility for Social Protection Programs

Ana Areias and Matthew Wai-Poi

Introduction

This chapter presents a nontechnical summary of the main findings of Areias, Keleher, Ralston, and Wai-Poi ([Areias et al.] forthcoming). In the paper, which was commissioned for this book, the authors note that technological advances may offer new opportunities for assessing beneficiary eligibility for social protection programs. In particular, they examine the following:

- *New techniques:* machine learning applied to existing household survey data to improve proxy means testing (PMT) models
- *New data:* machine learning applied to new data sources to improve targeting outcomes.

Chapter 6 surveys the different ways in which new data—such as remote sensing data, call data records, and administrative records, often made usable through machine learning processing—can make geographic targeting easier, more accurate, cheaper, or more frequent.

This current chapter reviews the latest methods in machine learning and their application to household welfare prediction and compares the results with those from traditional PMT models used for targeting social assistance programs. The chapter is designed for nontechnical readers.

507

It summarizes the results of a comprehensive, multicountry simulation of different PMT methods, including machine learning approaches, applied to around 20 Sub-Saharan African country-years. A technical summary of the different models and references to key texts is included in Areias et al. (forthcoming).

Areias et al. (forthcoming) conclude that machine learning applied to existing household surveys does not significantly improve predictive performance. A comprehensive and systematic assessment of both traditional and new machine learning models is conducted with the standard household survey data used in traditional PMT modeling. The paper finds that the results depend on the following:

- Which policy maker objective is being optimized (for example, inclusion or exclusion errors)
- Whether the same scoring system is being used across programs of different sizes
- How the scoring system is implemented (for example, by the line or quota method).

Moreover, even when particular models dominate in certain contexts, the magnitude of the improved performance is limited in a practical sense.

Nevertheless, Areias et al. (forthcoming) note the limitations of the particular experiment conducted and the conditions in which machine learning offers greater promise for improving the accuracy of social protection eligibility assessments. Although this chapter summarizes a systematic assessment of machine learning algorithms, it is only a static assessment. It applies different machine learning algorithms to the standard household cross-sectional survey data used by traditional PMT models to predict whether a household is below a particular income or consumption threshold. It does not allow for continuous recalibration of models based on dynamic data collection, for which machine learning approaches may offer significantly more promise as more data are available (see, for example, Noriega-Campero et al. 2020; Russom 2018). Further, the chapter does not attempt feature engineering, which has been shown in some settings to improve targeting outcomes and does not incorporate nonsurvey data sources, such as administrative data, which could also improve model accuracy. The potential for these approaches is discussed in chapter 6.

Practical Considerations for Policy Makers

New techniques and new data hold the promise of helping policy makers predict household welfare more accurately. Recent papers (reviewed in chapter 6) explore the use of sensing data and machine learning techniques, often in combination with traditional data but not always. Studies have

found a good correlation with welfare measures at a given geographical level, but the results are still inconclusive for individual or household targeting. Hence, Areias et al. (forthcoming) assess whether PMT models for household welfare prediction can be improved with new machine learning algorithms.

In assessing these models, it is important to consider the practical context in which such assessments take place. Some techniques or data may be better suited to evaluating welfare at a certain level (such as spatially disaggregated small areas) or for a certain purpose (such as identifying areas or households subject to natural shocks and agricultural risks), rather than more standard household monetary welfare assessments, which generate specific estimates for each household. Moreover, many countries have a range of different targeted social protection programs, which may differ in terms of size and target beneficiaries. These differences may influence the appeal of the new approaches, and such considerations also apply when considering whether PMT models for household poverty prediction are improved with new machine learning algorithms. These practical and policy considerations also pertain to traditional PMT models, but the complexity of the new machine learning approaches compounds them. For example, generating multiple scoring models for different cutoff points or parts of the country requires considerably more time and effort to develop and can be difficult to communicate to policy makers and the public. Whether any degree of improved accuracy warrants these complications is a trade-off to be assessed. In Indonesia, using different scoring models for each social protection program and their different eligibility thresholds was considered. The approach was rejected by the government and World Bank technical team as being too complicated to communicate to nontechnical policy makers (such as in the line ministries implementing the programs). Moreover, even a single traditional PMT model is often subject to the criticism that it is a "black box," both because the statistical approach can be difficult for a lay audience to understand and because the exact scoring formula is often kept confidential to avoid households gaming the scoring system. Explaining the more complex machine learning algorithms, such as computational neural networks, is strikingly more difficult.

Therefore, the conclusion on whether to incorporate the new models depends not just on whether they show systematic improvements over existing PMT approaches, but also on practical considerations in a targeting context.

Assessing Machine Learning for Household Welfare Prediction: Potential and Provisos

The appeal of machine learning methods for prediction of household welfare is twofold (Athey 2018). First, even outside machine learning models

themselves, common machine learning processes can help improve traditional PMT models to prevent overfitting, where the models are very good at predicting the data used to develop them but not the real-world data used to predict household welfare and determine eligibility, as is discussed in the next section. Second, and most importantly, machine learning algorithms can discover complex relationships across many dimensions of data, suggesting that they might be able to predict household welfare more accurately.

However, there is an important limitation to the analysis in this chapter. Although the authors believe that this is the largest systematic assessment in terms of the number of machine learning algorithms included and the number of country data sets to which they are applied, the data sets are all for African countries (and in some cases, they are for the same country in different years). Thus, the conclusions about which algorithms perform the best in different circumstances may not generalize to other country and regional contexts, particularly those with higher levels of income or at a different range of potential survey variables on which to model. However, the variation in the results, which depend on both the manner in which scoring is implemented and the policy maker's objectives—whether they favor minimizing inclusion error, minimizing exclusion error, or a balance of both—warrants caution in the application of machine learning to household welfare prediction in other regions as well.

Applying Machine Learning Processes to Traditional PMT Models

Even before offering the potential of better modeling, machine learning processes are already improving traditional PMT approaches. Before turning to the main assessment of the different models and their prediction performance, this section describes two important machine learning–driven processes that are already transforming the performance of otherwise traditional PMT models. The first is an improved method for variable selection. The second is an improved method for evaluating ex ante model performance. This discussion builds on that in chapter 5.

Variable Selection

Determining which and how many household and neighborhood characteristics should be used in models has a critical influence on their performance. In chapter 5, the discussion on variable selection notes the danger of overfitting, where the model is very good at predicting the survey data on which the model is based but not very good at predicting new data from

outside the model, which are collected from the households on which it will be applied. Should an analyst include all available variables from the survey data and see what the model prefers? Or should the analyst screen out some variables ahead of the main modeling? Without analyzing every possible combination (*best subset selection*), which is computationally burdensome, some process is needed to decide which variables to include in the predictive model. The common approaches for PMT models have historically been to include all the variables and keep those that are statistically significant, or to preselect variables through a stepwise approach that introduces or eliminates variables one at a time and uses a statistical test to determine whether the model fit is improved. However, as chapter 5 notes, this approach may exclude the optimal combination of variables.

A machine learning approach to variable selection, called penalized regression, offers a superior approach to variable selection than traditional PMT methods. A standard method for variable selection in machine learning models is to include a penalty for complexity. This results in simpler models with fewer but the most predictive variables and tends to improve how well the models perform when evaluating data outside those with which they were developed. Common machine learning penalized algorithms include Lasso and Ridge regressions, or a combination of the two called elastic net. Intuitively, these models ask whether the inclusion of a variable improves prediction enough to warrant inclusion, but unlike the stepwise approach (which does the same), it is not dependent on the order in which variables are evaluated. As chapter 5 notes, this approach has already been used in several countries.

Model Validation

Traditional approaches to validating PMT models involve splitting the survey data into two samples, the first for training the model and the second for testing how well it performs on nontraining data.[1] In the simplest (but most naive) approach, the analyst runs the PMT regressions of consumption or income on the proxy variables using the entire household survey. The analyst then assesses how well the model performs—for example, by examining inclusion and exclusion errors—on the same data on which the model was developed. More appropriately, the analyst splits the data into training and test data sets. The model is developed on the training data and then its performance is assessed on the test data. In this way, model performance is assessed on data that are new to the model, not on the same data used to develop it. Commonly, the PMT analyst will split the data on a 1:1 or 2:1 basis.

Machine learning also splits data into training and test data sets, but it uses an approach that ensures that every observation is used both to train and to

test, improving model validation. A standard machine learning approach (called *k*-fold cross validation) splits observations into multiple groups and takes turns estimating the models on some of the groups and testing them on other groups. For example, the data might be split into 10 random and equal groups (or folds). The model is estimated using nine of the groups as the training set and the last group as the test set. This process is then repeated after moving the test group into the training set and swapping one of the previous training groups out to be the new test set. Once all the groups have been in both the training and test sets, the model performance is assessed by averaging the errors across all iterations. This has the advantage over just using a single split of ensuring that every observation can appear in both the training and test sets.[2] Thus, an important contribution that machine learning already provides to traditional PMT modeling is a data-driven approach that helps avoid overfitting to a greater extent than traditional approaches.

Machine Learning Models

Beyond standard processes, various methods in the machine learning literature could potentially improve model accuracy when used to predict the same outcome variable with the same explanatory variables as in traditional PMT models. In this nontechnical section, the model categories, philosophies, and flavors are not elaborated, they are just listed. A fuller treatment is provided in Areias et al. (forthcoming), based on James et al. (2013) and Kuhn and Johnson (2018). The following model categories are examined:

- Linear models[3]
- Robust models[4]
- Penalized regressions[5]
- Nonlinear models[6]
- Tree-based models.[7]

Different models have different characteristics that make them more or less attractive on different dimensions. Areias et al. (forthcoming) and supporting references discuss these characteristics in more detail, but they include the following:

1. Can they be used when there are more characteristics than there are observations ($n < p$)?
2. Do they require significant data cleaning before modeling?
3. Are the results easily interpreted?
4. Do they automatically select variables or does the analyst have to do this?
5. Do they require tuning: are there parameters on which the model must optimize or the user must select (or both), indicating a degree of model complexity and processing intensity?

6. Are they robust to noisy data: can the algorithm deal with noise in the data that is not related to household welfare?
7. Are they computationally intensive?

No one algorithm dominates (see table 8B.2). As Kuhn and Johnson (2018, 25–26) observe, "[i]t is our experience that some modeling practitioners have a favorite model that is relied on indiscriminately," while noting Wolpert's (1996) "No Free Lunch" theorem, "which argues that without having substantive information about the modeling problem, there is no single model that will always do better than any other model." They conclude that, generally, "no one model is uniformly better than the others. The applicability of a technique is dependent on the type of data being analyzed, the needs of the modeler, and the context of how the model will be used" (Kuhn and Johnson 2018, 549). For the purposes of predictive models of household welfare and social protection applications, particularly of note is that the traditional PMT models (linear and logistic regressions) are easily interpretable and computationally easy, with no tuning parameters. At the same time, they require significant preprocessing, have no automatic variable selection, and are not robust to predictor noise.

Assessing Machine Learning Performance on PMT Data: A Benchmarking Experiment

Setting Up the Experiment

The Areias et al. (forthcoming) benchmarking experiment is an exhaustive comparison of models over a wide range of African data sets, including the standard variables used in PMT models. The exercise, which assesses traditional PMT and alternative machine learning algorithms, consists of the following: (1) 17 country data sets, (2) 19 algorithms, (3) four performance measures, and (4) 100 learning and validation samples.[8] To evaluate the results, a mixed-effects methodology is applied. The exercise uses 17 harmonized African data sets, the country characteristics of which are included in annex 8A. The countries are all low-income or lower-middle-income countries, except Mauritius; they range in population from fewer than 1 million (the Comoros) to nearly 20 million (Niger); their populations are mostly very young (40–50 percent of the population is younger than age 15 years); and they are rural (all but two are less than 50 percent urbanized), uneducated (in most of the countries, the majority have not completed primary education), and poor (in only two countries is the $1.90 poverty rate below 25 percent).[9] For each data set, 100 random resamples were drawn with a range of variables including the following:

- *Location:* region, subregion, urban/rural, and capital city
- *Demographics:* household size (number of children, males, females, and elderly of various ages) and the household head's sex, age, marital status, and education
- *Dwelling:* ownership; number of rooms; material of the roof, walls, and floor; source of water and electricity; and type of toilet, cooking and lighting fuel, and garbage disposal
- *Durables:* radio, television, land phone, cell phone, refrigerator, sewing machine, computer, stove, animal cart, bicycle, boat, motorcycle, car, access to internet, large livestock, medium livestock, and poultry
- *Land ownership:* agricultural land and nonagricultural land
- *Employment:* employment status and labor force activity, type and sector of employment, wages, and hours worked
- *Distance and time to facilities:* water, school, and health center.

In addition, the analysis summarized in this chapter incorporates the survey design, accounting for survey strata and household weights. Nineteen different algorithms are assessed, which include both traditional PMT models and newer machine learning models.[10]

Implementing the Scoring Models

There are two different practices in the field for implementing scoring models. The first uses a *threshold* eligibility approach. In this case, program eligibility is set as all those households with consumption below a certain line. In some countries and programs, this means the official national poverty line; in others, it may be a higher or lower line depending on the budget and program objectives. Households with PMT scores below this line become beneficiaries, regardless of how many or few beneficiaries there are as a result. For the simulations, the national poverty line is used for each data set, which means that this approach determines algorithm performance over different poverty lines and rates across countries.

The second practice uses a *quota* approach, in which the number of program beneficiaries is fixed (based on the budget or the estimated number of poor and vulnerable). Households then become beneficiaries starting with those with the lowest PMT scores first and continuing until the quota is met. For the simulations, Areias et al. (forthcoming) use a constant 10 percent quota of households for all the data sets, which means that this approach determines algorithm performance for a fixed program coverage across different countries.[11]

Assessing Model Performance

A range of performance measures are used to evaluate the set of models. Ideal measures for targeting performance are difficult to construct,

as chapter 6 examines in depth. The Areias et al. (forthcoming) analysis uses four measures from the machine learning literature to assess model performance. These are discussed in more detail in chapter 7, but they can broadly be understood as assessing the following objectives:

- Normalized root mean square error (NRMSE): a standard machine learning and regression performance measure that evaluates how well the algorithm predicts across the entire income/consumption distribution[12]
- F_1: a measure that intuitively balances inclusion and exclusion errors, weighting each equally
- F_2: a measure of inclusion and exclusion errors that weights exclusion errors twice as much as inclusion errors
- Matthews correlation coefficient (MCC): a balanced measure of inclusion and exclusion errors that is invariant to the size and scale of programs and poverty rates.[13]

Areias et al. (forthcoming) assess predictive performance on all four measures for two reasons. First, NRMSE is the standard metric used in machine learning analysis to evaluate regression model performance in general. It is not clear ex ante that this will necessarily result in the best targeting performance, since it is assessing how well a model predicts the entire distribution and not around particular sections of interest, such as the poor. Second, the three targeting-related errors allow for assessing how well a model performs relative to policy makers' different objectives. F_1 reflects some policy makers' preference to consider both inclusion and exclusion errors as equally concerning. F_2 reflects other policy makers' preference to avoid excluding the poor by mistake, even if this means including more nonpoor as beneficiaries. MCC considers both inclusion and exclusion errors equally, but it is calculated in a way that it is comparable across programs of different sizes or countries with different poverty rates or targets. While Areias et al. (forthcoming) and chapter 7 provide a technical discussion of these measures, F_2 (which places greater emphasis on avoiding exclusion of the poor) is perhaps most aligned with the distributional characteristic favored in chapter 7, which weights the impact on the poor the most.[14]

Given the very large number of simulations, the entire set of outcomes is itself subjected to a statistical assessment. Although it is easy to present the results of a few analyses visually, the large numbers of algorithms, data sets, and resamples ($19 \times 17 \times 100 = 32,300$ run with four performance measures each) mean that the results are more easily summarized using a mixed statistical model for a meta-analysis, the methodology for which is discussed in annex 8B.[15]

Main Results

To understand the results of the systematic assessments, several questions are examined step-by-step:

- Are there statistically significant differences in model performance across all the countries? Are these differences large enough to care?
- Does the choice of the measure of targeting performance matter? In other words, does a policy maker's objective (prioritizing exclusion error versus balancing exclusion and inclusion errors) change how the models are assessed?
- Does the choice of targeting implementation matter? Is there a difference when the same model is used with the threshold approach compared with the quota approach?
- In the case that a small number of models have superior performance, what else should be considered when selecting between them?

Overall Model Choice May Not Matter: There Are Few Differences between Models, Especially Sizable Differences, When They Are Assessed Using Standard Machine Learning Metrics

There are different ways of thinking about algorithm performance. First, one model's performance can be compared with another's to see if one is superior in a statistically significant sense. The results of such comparisons are called "significant" wins and losses. The assessment could also ask whether these differences, although statistically significant, are large enough to matter in a practical sense. These are called "relevant" wins or losses, defined here as the difference in model performance being statistically significant and greater than 5 percent. Second, the assessment can compare one model against all the other models and ask how many it is statistically better or worse than. If a model's performance is statistically no different from a number of others, they are in the same equivalence class.

When one model is compared one at a time with each of the others, there are few models that are frequently statistically better or worse, and almost no differences are substantive. The two bars in figure 8.1 show the net significant and relevant wins or losses for each model when compared with another and using the NRMSE performance measure. The most significant wins for any model are six (of 18; the logistic model does not produce the root mean square error), and almost all have two or fewer wins. Two models in the robust class have many losses, but this is largely a technical result.[16] Even more tellingly, when the significant wins are restricted to also being relevant, almost all the models have zero wins; only

three have one win each. Moreover, when each model is compared with all the others, the 12 best performing models cannot be statistically distinguished from each other. That is, although one of the tree models has 6 net wins, it is statistically not better than 11 other algorithms when they are considered as a group. Annex 8B shows the relative performance of each model against all the others simultaneously.[17]

Figure 8.1 Pairwise Wins Using the Standard Machine Learning NRMSE Performance Measures
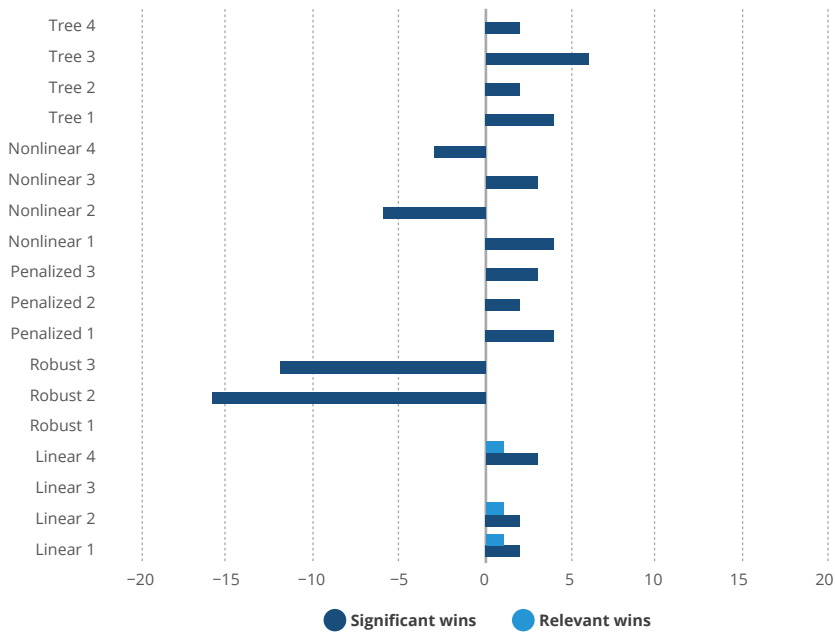
Source: Areias et al., forthcoming.

Note: Linear 1 = OLS without variable selection; Linear 2 = OLS with stepwise selection and Akaike Information Criterion (AIC); Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression with stepwise AIC variable section; Robust 3 = quantile regression with cross-validated quantile and stepwise AIC variable section; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees. AIC = Akaike Information Criterion; NRMSE = normalized root mean square error; OLS = ordinary least squares.

Note: Significant wins means one algorithm performed better than another in a statistically significant manner. Relevant wins means not only did it perform better in a statistical manner but the difference in performance was relevant in a practical sense, used here to mean a 5 percent difference or greater.

Policy Choice Matters: Performance on Standard Machine Learning Metrics Is Different from Performance on Targeting-Related Metrics

Performance on standard machine learning metrics does not necessarily equate to performance on targeting metrics. The results examined in the previous section relate to how well models predict the entire household welfare (income/consumption) distribution according to NRMSE. However, the targeting analyst's role is to categorize households as eligible or not eligible. This does not require getting the shape of the distribution right, or even getting the rank order of households right. It just requires correctly estimating each household as being above or below the program eligibility threshold, alternatively interpreted as having a PMT score below the program income/consumption threshold (simulated here as the national poverty line approach) or among the poorest households ranked by PMT score up to the program quota (simulated here as the poorest 10 percent approach). Algorithms that may not perform well across the entire distribution might nonetheless be superior when attempting to classify program eligibility, especially at the poorer end of the distribution.

Indeed, some models that perform well on standard machine learning measures do not perform well on targeting-related measures. This section looks at performance on the three targeting measures using the field practice of targeting all households with a PMT score indicating they are below the national poverty line (which differs across data sets due both to how the poverty line is set and the national income/consumption distribution). For all the metrics, a higher score is better.[18] All the model results are presented relative to the baseline, which is a standard ordinary least squares (OLS) model without variable selection.[19] The first result to note is that optimizing models for NRMSE (a standard machine learning approach) does not necessarily optimize models for targeting-related measures. Figure 8.2 compares outcomes for each algorithm relative to a baseline OLS regression with no variable selection (the simplest traditional PMT approach), using the NRMSE and F_2 measures (F_2 is the targeting measure that over-weights exclusion error).[20] One algorithm that performs poorly on NRMSE, *Robust 3*, is one of the best relative performers on the F_2 measure.

Moreover, model performance also varies depending on the targeting objective of the policy maker with respect to inclusion and exclusion errors. Figure 8.3 shows that some models perform quite differently on the different targeting metrics. *Robust 3* underperforms on the MCC measure (scale-invariant and balances inclusion and exclusion errors) but outperforms on the F_1 measure (balances inclusion and exclusion errors

**Figure 8.2   NRMSE and F_2 Estimated Coefficients Relative to the
Baseline, Poverty Approach**

● F_2 **(analogous to inclusion error x1 and exclusion error x2)**

● **NRMSE (standard machine learning measure for distribution, inverted)**

Source: Areias et al., forthcoming.

Note: 95 percent confidence intervals for algorithm performance on each measure relative to a reference algorithm Linear 1.

Note: Linear 2 = OLS with stepwise selection and AIC criterion; Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression with stepwise AIC variable section; Robust 3 = quantile regression with cross-validated quantile and stepwise AIC variable section; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees. AIC = Akaike Information Criterion; NRMSE = normalized root mean square error; OLS = ordinary least squares.

but depends on program size) and, particularly, the F_2 measure (over-weights exclusion error and depends on program size). That is, if the policy maker cares about balanced targeting performance across a range of program sizes, this may not be the best model. If the policy maker cares about exclusion error for a particular program size, this may be a better model. Moreover, four models (*Linear 5*,[21] *Robust 3*, *Tree 2*, and *Tree 4*) show much better relative performance on the F_2 measure (which prefers lower exclusion error) than on the other two measures.[22] Nonetheless, around two-thirds of the confidence intervals have no statistical

Figure 8.3 **Matthews Correlation Coefficient, F_1, and F_2 Estimated Coefficients Relative to the Baseline, Poverty Line Approach**

- **Matthews correlation coefficient (analogous to balanced inclusion error and exclusion error, scale-invariant)**
- **F_1 (analogous to balanced inclusion error and exclusion error)**
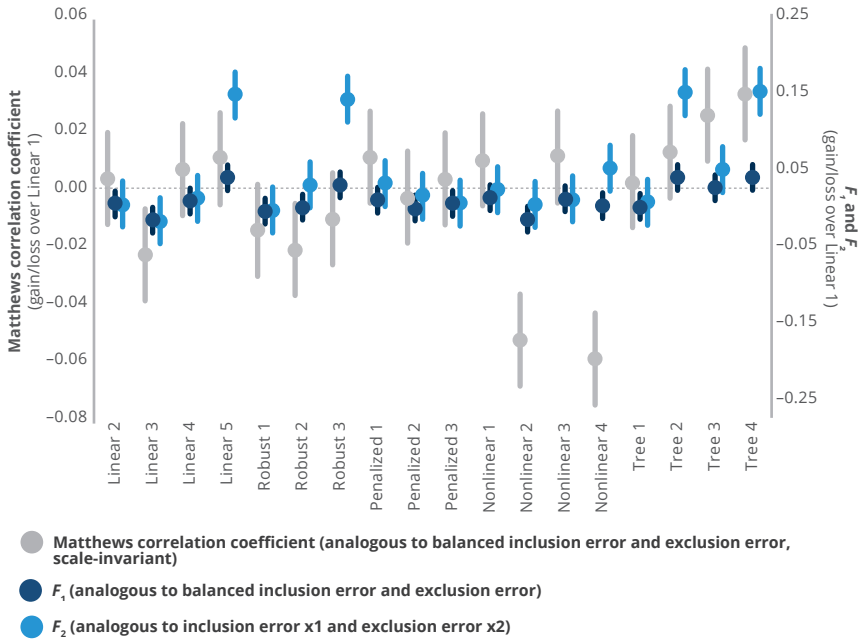- **F_2 (analogous to inclusion error x1 and exclusion error x2)**

Source: Areias et al., forthcoming.

Note: 95 percent confidence intervals for algorithm performance on each measure relative to a reference algorithm Linear 1.

Note: Linear 2 = OLS with stepwise selection and AIC criterion; Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Linear 5 = logistic with Lasso variable selection; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression with stepwise AIC variable section; Robust 3 = quantile regression with cross-validated quantile and stepwise AIC variable section; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees. AIC = Akaike Information Criterion; OLS = ordinary least squares.

difference from the baseline, indicating that most of the algorithms do not outperform the most basic PMT algorithm on the various targeting measures. Together, these two results highlight that the policy maker's objectives matter: some models likely do better at modeling the entire distribution rather than targeting the lower end,[23] or do better at minimizing exclusion error relative to a more balanced targeting measure or a measure that is scale invariant and therefore more comparable across different programs and periods.

Model Performance Differs by Which Metric Is Used but Is Generally Irrelevant When the Magnitude of the Difference Is Considered

Table 8.1 summarizes the net statistical and relevant pairwise wins across all the measures, including the sum of the three machine learning measures. The results from the relative estimated coefficients are also evident here. Some algorithms perform poorly on NRMSE (standard machine

Table 8.1 Significant and Relevant Pairwise Wins, Poverty Line Approach

Algorithm	NRMSE	MCC	F_1	F_2	All 3 ML	NRMSE	MCC	F_1	F_2	All 3 ML
Linear 1	2	1	−3	−4	−6	1	0	0	−4	−4
Linear 2	2	1	−3	−4	−6	1	0	0	−4	−4
Linear 3	0	−6	−5	−6	−17	0	−1	0	−4	−5
Linear 4	3	3	−3	−4	−4	1	0	0	−4	−4
Linear 5	NA	4	11	15	30	NA	0	0	4	4
Robust 1	0	0	−5	−5	−10	0	0	0	−3	−3
Robust 2	−16	−5	−3	−4	−12	0	−1	0	−3	−4
Robust 3	−12	0	4	15	19	0	0	0	6	6
Penalized 1	4	4	0	−4	0	0	0	0	−2	−2
Penalized 2	2	1	−4	−4	−7	0	0	0	−2	−2
Penalized 3	3	1	−3	−4	−6	0	0	0	−2	−2
Nonlinear 1	4	4	0	−4	0	0	0	0	−2	−2
Nonlinear 2	−6	−17	−5	−4	−26	0	−5	0	−2	−7
Nonlinear 3	3	4	0	−4	0	0	1	0	−2	−1
Nonlinear 4	−3	−17	−3	−2	−22	0	−4	0	−2	−6
Tree 1	4	1	−3	−4	−6	0	2	0	−2	0
Tree 2	2	4	11	15	30	0	2	0	14	16
Tree 3	6	6	3	−3	6	0	2	0	−1	1
Tree 4	2	11	11	15	37	0	4	0	15	19

Source: Areias et al., forthcoming.

Note: Linear 1 = OLS without variable selection; Linear 2 = OLS with stepwise selection and Akaike Information Criterion; Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Linear 5 = logistic Lasso with variable selection; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression; Robust 3 = quantile regression with cross-validated quantile; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees. MCC = Matthews correlation coefficient; ML = the sum of the three machine learning measures of MCC, F_1, and F_2; NA = not applicable; NRMSE = normalized root mean square error; OLS = ordinary least squares.

learning measure for the entire distribution) compared with targeting measures (for example, *Robust 2* and *Robust 3*). Some perform better on the F measures (which depend on program size) than scale-invariant MCC (for example, *Linear 5* and *Tree 2*) and many even have opposite performance (*Linear 1*, *Linear 2*, *Linear 4*, *Penalized 1*, *Nonlinear 1*, *Nonlinear 3*, and *Tree 1*). However, as the relevant win section of the table shows, very few of these results matter in a practical sense, with the magnitude of the difference being less than 5 percent even when statistically significant.
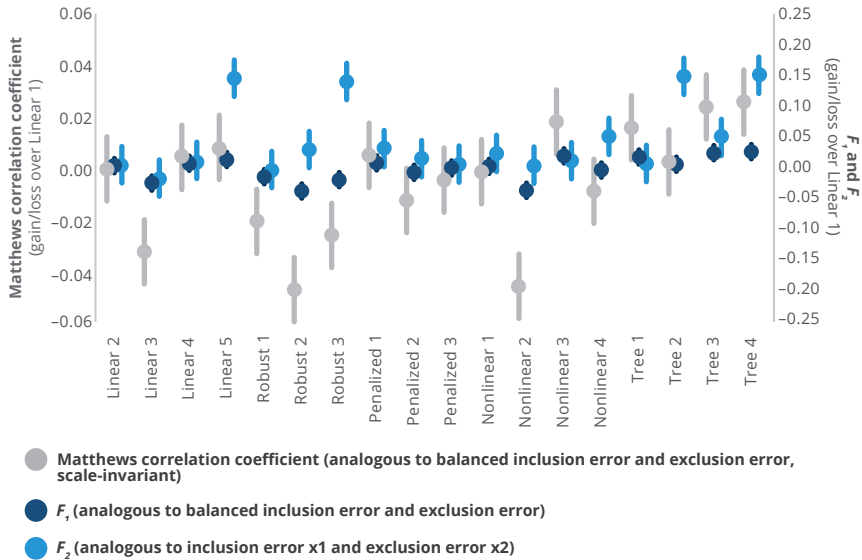
When favoring exclusion error, some models are clearly preferred or not preferred for these particular country data sets. Considering the F_2 score, which pays more attention to exclusion error, many more differences are relevant. Two algorithms dominate (*Tree 2* and *Tree 4*) and two (*Nonlinear 2* and *Nonlinear 4*) tend to perform relatively poorly on all three targeting measures.

Implementation Matters: Outcomes Using a Threshold Approach Are Different from Outcomes Using a Quota Approach

Although different models perform better under different metrics using the quota approach, it is the different models that demonstrate this. Next, the alternative field practice of ranking households by PMT score and selecting those with the lowest until the poverty quota is met is examined. As with the poverty line approach, relative performance can vary for the same algorithms on different targeting measures, with some significantly positive performers on F_2 (scale dependent and over-weighting exclusion error) compared with their MCC and F_1 performances (both balance inclusion and exclusion error, but F_1 is scale dependent). *Linear 5* is similar to the baseline on the F_1 and MCC measures, but it is significantly better for F_2, while *Nonlinear 4* is slightly worse than the baseline on MCC and F_1, but it is better on F_2 (figure 8.4). However, while these results are similar to those from the poverty line approach—some algorithms perform differently on different metrics—the algorithms demonstrating this under the poverty quota approach are different from those under the poverty line approach, except for *Linear 5*.

Some models that do better than the baseline using a threshold approach do worse than the baseline using a quota approach. Under the threshold approach, *Robust 3* outperforms the baseline on the F_1 (scale dependent, balances inclusion and exclusion errors) and particularly F_2 (scale dependent, over-weights exclusion error) measures; under the quota approach, *Robust 3* underperforms the baseline on all measures. Figure 8.5 shows the results for all the algorithms on the three targeting measures under the quota and threshold approaches. Algorithms with the same

Figure 8.4 Matthews Correlation Coefficient, F_1, and F_2 Estimated Relative to the Baseline, Poverty Quota Approach

● Matthews correlation coefficient (analogous to balanced inclusion error and exclusion error, scale-invariant)

● F_1 (analogous to balanced inclusion error and exclusion error)

● F_2 (analogous to inclusion error x1 and exclusion error x2)

Source: Areias et al., forthcoming.

Note: Linear 2 = OLS with stepwise selection and AIC criterion; Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Linear 5 = logistic with Lasso variable selection; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression with stepwise AIC variable section; Robust 3 = quantile regression with cross-validated quantile and stepwise AIC variable section; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees. AIC = Akaike Information Criterion; OLS = ordinary least squares.

performance under both approaches lie on the 45-degree line, and those in the upper right or the lower left quadrants indicate a similar number of net wins and losses, regardless of whether the quota or threshold approach is used. Several points lie near the 45-degree line and many more are in the upper right or lower left quadrants. However, within these two quadrants, several points are far from the 45-degree line, signifying that the number of net wins or net losses varies considerably depending on the targeting implementation approach. Furthermore, particularly on the F_1 and F_2 measures (scale dependent, the former balanced and the latter over-weighting exclusion error), many algorithms are likely to have net losses under the threshold approach but net wins under the quota approach (bottom right) or vice versa (top left). This combination of a similar result driven by different

Figure 8.5 **Net Significant Wins by the Poverty Quota and Poverty Line and Approaches, by Targeting Measure**
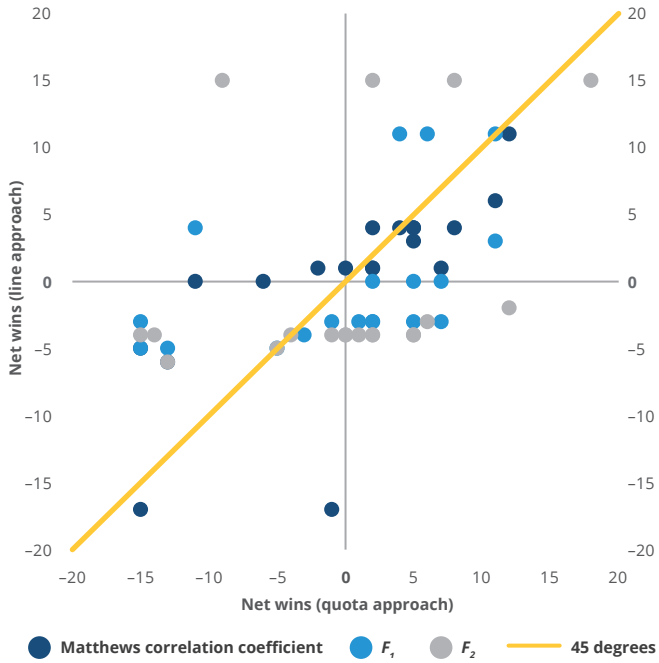
Source: Areias et al., forthcoming.

algorithms or a different result for the same algorithm over the two approaches highlights that it is not just a policy maker's targeting objective (in terms of targeting errors to minimize) that matters, it is also the mode of implementation.[24]

Taken together, the results show that model performance depends on both policy makers' objectives and how the scoring is implemented in practice, but also that the differences in outcomes are generally not large enough to matter. Table 8.2 provides a full summary of the results when each algorithm is applied using both the line and quota approaches, in terms of significant wins (panel A) and relevant wins (panel B). The results on machine learning applied to PMT data discussed thus far are visually summarized in this table.

1. An algorithm's performance depends on the targeting measure it is optimizing and being used to evaluate it (which is itself the policy maker's choice). For example, in panel A under the line approach, *Penalized 1*,

Table 8.2   Summary of Algorithm Rankings by Targeting Measure and Poverty Line or Quota Approach

Panel A: Significant wins

	Poverty line					Poverty quota				
Algorithm	NRMSE	MCC	F_1	F_2	All 3 ML	NRMSE	MCC	F_1	F_2	All 3 ML
Linear 1	2	1	−3	−4	−6	2	2	2	0	4
Linear 2	2	1	−3	−4	−6	2	2	2	1	5
Linear 3	0	−6	−5	−6	−17	0	−13	−13	−13	−39
Linear 4	3	3	−3	−4	−4	3	5	5	2	12
Linear 5	NA	4	11	15	30	NA	5	6	18	29
Robust 1	0	0	−5	−5	−10	0	−6	−5	−5	−16
Robust 2	−16	−5	−3	−4	−12	−16	−15	−15	−15	−45
Robust 3	−12	0	4	15	19	−12	−11	−11	−9	−31
Penalized 1	4	4	0	−4	0	4	5	5	2	12
Penalized 2	2	1	−4	−4	−7	2	−2	−3	−4	−9
Penalized 3	3	1	−3	−4	−6	3	0	1	−1	0
Nonlinear 1	4	4	0	−4	0	4	2	2	0	4
Nonlinear 2	−6	−17	−5	−4	−26	−6	−15	−15	−14	−44
Nonlinear 3	3	4	0	−4	0	3	8	7	5	20
Nonlinear 4	−3	−17	−3	−2	−22	−3	−1	−1	12	10
Tree 1	4	1	−3	−4	−6	4	7	7	5	19
Tree 2	2	4	11	15	30	2	4	4	2	10
Tree 3	6	6	3	−3	6	6	11	11	6	28
Tree 4	2	11	11	15	37	2	12	11	8	31

Panel B: Relevant wins

	Poverty line					Poverty quota				
Algorithm	NRMSE	MCC	F_1	F_2	All 3 ML	NRMSE	MCC	F_1	F_2	All 3 ML
Linear 1	1	0	0	−4	−4	1	0	0	−1	−1
Linear 2	1	0	0	−4	−4	1	0	0	−1	−1
Linear 3	0	−1	0	−4	−5	0	−7	−5	−6	−18
Linear 4	1	0	0	−4	−4	1	1	0	−1	0
Linear 5	NA	0	0	4	4	NA	1	1	4	6
Robust 1	0	0	0	−3	−3	0	−3	−2	−2	−7
Robust 2	0	−1	0	−3	−4	−3	−9	−7	−7	−23
Robust 3	0	0	0	6	6	0	−4	−3	−3	−10
Penalized 1	0	0	0	−2	−2	0	2	1	1	4
Penalized 2	0	0	0	−2	−2	0	−1	0	0	−1
Penalized 3	0	0	0	−2	−2	0	1	0	0	1
Nonlinear 1	0	0	0	−2	−2	0	1	1	0	2
Nonlinear 2	0	−5	0	−2	−7	0	−6	−5	−6	−17
Nonlinear 3	0	1	0	−2	−1	0	5	4	3	12
Nonlinear 4	0	−4	0	−2	−6	0	2	0	5	7
Tree 1	0	2	0	−2	0	0	4	3	3	10
Tree 2	0	2	0	14	16	0	2	2	2	6
Tree 3	0	2	0	−1	1	0	5	5	4	14
Tree 4	0	4	0	15	19	0	6	5	5	16

—

Source: Areias et al., forthcoming.

Note: Linear 1 = OLS without variable selection; Linear 2 = OLS with stepwise selection and Akaike Information Criterion; Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Linear 5 = logistic Lasso with variable selection; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression; Robust 3 = quantile regression with cross-validated quantile; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees.
MCC = Matthews correlation coefficient; ML = the sum of the three machine learning measures of MCC, F_1, and F_2; NA = not applicable; NRMSE = normalized root mean square error; OLS = ordinary least squares.

Nonlinear 1, and *Nonlinear 3* perform above average on the NRMSE and MCC measures in terms of net significant wins, but they perform below average on the F_2 measure.

2. Algorithm performance also depends on the mode of targeting implementation—poverty line versus poverty quota (also the policy maker's decision). Looking at the results under the line and quota approaches in panel A, *Robust 3* is one of the best performers on the F_1 and F_2 measures when the line approach is used, but it is one of the worst when the quota approach is used. To a lesser extent, the converse is true of some of the other algorithms, such as *Nonlinear 4*.

3. Even when there are clear statistical winners, generally the differences in performance are not important in magnitude. Most of the results disappear when the threshold for a win is not statistically different (panel A) but more than 5 percent different on the targeting measure (panel B). This is less the case for F_2 in the case of the line approach and many measures in the case of the quota approach, but the pattern still holds.

4. The context clearly matters for an algorithm's relative performance: choice of measure and choice of implementation. Nonetheless, even when considering relevant wins only, some algorithms have consistently better or worse results across all contexts. Gradient boosted quantile regression trees (*Tree 4*) tend to outperform most of the other algorithms on all measures under both the line and quota approaches. The traditional logistic regression implemented with a simple Lasso variable selection model (*Linear 5*) also provides no reason not to use it in any of these contexts (although on relevant wins the differences are marginal). The *k*-nearest neighbors model (*Nonlinear 2*) is clearly an inferior choice across all measures and modes of implementation, even considering only relevant wins.

5. In addition, even when the conclusions are limited to the current data sets, the choice between the best performing gradient boosted quantile

regression trees (*Tree 4*) and the reasonably performing traditional logistic regression implemented with a simple Lasso variable selection model (*Linear 5*) may depend also on other considerations. These two algorithms have almost opposite characteristics (see Kuhn and Johnson 2018, table A.1). Boosted models allow for $n < p$ (more predictors than observations, meaning a greater range of models can be explored), whereas the logistic models do not. Logistic models also require significant preprocessing and are not robust to predictor noise, while boosted models do not require preprocessing and are robust.[25] However, logistic models are easier to interpret than boosted models, do not require optimization or choice of tuning parameters (boosted models require three or four choices), and are computationally much less intensive.[26]

The following list summarizes the results and conclusions from the Areias et al. (forthcoming) systematic assessment of traditional and machine learning algorithms for PMT from 17 African data sets:

- Result 1. There is generally more variation in algorithm performance within models across different data sets than between models across all data sets.
- Result 2. No algorithm dominates in all circumstances.
- Result 3. Although several performance differences are statistically significant, there are many fewer at a more meaningful level (at least a 5 percent difference in performance). Most of the machine learning results do not outperform the standard PMT approach.
- Result 4. The standard machine learning model optimization (minimizing the root mean square error) does not produce the same results as optimizing for targeting-related performance measures. The models that are best at predicting outcomes across the distribution may not minimize targeting errors.
- Result 5. The results vary across different targeting measures (and hence objectives). Some algorithms do not perform well on balanced targeting measures but outperform on measures that aim to reduce exclusion error.
- Result 6. For selected purposes (for example, focusing on reducing exclusion error), a few algorithms dominate, and one is categorically worse than the others.
- Result 7. The results not only vary across targeting measures, but also the variation in results varies depending on how targeting is implemented; targeting objectives and practices affect performance.
- Conclusion 1. An algorithm's performance depends on the targeting measure it is optimizing and by which it is being measured—which are choices of the policy maker.
- Conclusion 2. Algorithm performance depends on the manner of targeting implementation.

- Conclusion 3. Even when there are clear statistical winners, generally the differences in performance are not important in magnitude.
- Conclusion 4. For the data sets examined here, some algorithms consistently perform better or worse than others. Notably, gradient boosted quantile regression trees tend to outperform most of the other algorithms under both implementation modalities. The traditional logistic regression with a simple Lasso variable selection model also proves robustly useful. The *k*-nearest neighbors model is clearly an inferior choice across all the performance measures and modes of implementation.
- Conclusion 5. Even when limiting the conclusions to the current data sets, the choice between gradient boosted quantile regression trees and the logistic model may depend on other considerations. The former allows for $n < p$ (more predictors than observations, meaning a greater range of models can be explored), whereas the logistic model does not. Logistic models also require significant preprocessing and are not robust to predictor noise, while boosted models do not require preprocessing and are robust. However, logistic models are easier to interpret than boosted models, do not require optimization or choice of tuning parameters (boosted models require three or four choices), and are computationally much less intensive. The particular implementation of the logistic model in Areias et al. (forthcoming) tries to split the difference, using Lasso for variable selection, which improves robustness and allows for $n > p$ but now does require optimization and tuning parameters.

Although the current analysis is applied to a specific set of African data sets, the results likely apply more widely. There may be concern that the current results may not generalize to non-African data sets. Nonetheless, the results still provide important considerations for the use of machine learning prediction in other settings. The results show that performance can vary across different dimensions, including the choice of performance measure, the policy maker's objective, program size, and the scoring implementation approach. Does the policy maker prioritize reducing exclusion error (analogous to the F_2 measure) or more balanced inclusion and exclusion errors (analogous to the F_1 measure)? Is the policy maker trying to use the same score to determine edibility for programs of very different sizes (in which case the MCC measure might be preferred)? Is the poverty line or quota approach being used? Regardless of whether the exact results of this work extend to other settings, they indicate that there is not necessarily a universally best algorithm and that if an optimal algorithm does exist for a particular context, context will matter. Whether an exhaustive search for the optimal algorithm is justified by the often limited real-world (if not statistical) differences between model outcomes is an important consideration for policy makers.

Moreover, this variation suggests a trade-off between multiple scoring models to optimize performance across different programs and complexity in practice. For example, if an analyst determines after exhaustive research that one algorithm is better for a particular program based on its size, objectives, and beneficiaries, and another algorithm is better for another program with different characteristics, should different scoring models be developed and applied for each program? What if there are three or four different programs and optimal models? The time requirements and difficulty of communicating model differences to policy makers meant that this approach was rejected in Indonesia with traditional linear models, although it was believed that different linear models could determine eligibility for larger and smaller programs with different degrees of accuracy. Once the complexity of machine learning algorithms is included in the consideration, model selection with differing performance across contexts becomes more difficult.

Finally, three areas stand out for future research. First, Areias et al. (forthcoming) analysis could be extended to other regions and income levels to see how generalizable the results are. Second, in a related sense, research could also focus on when different models or targeting contexts matter for prediction performance. For example, which is more important: (1) data sets (for example, the range and number of variables), (2) country context (for example, the degree of poverty and inequality), or (3) choice of algorithm? It is not even clear which methodology is best to answer this question. For example, a random forest model was applied to the tens of thousands of models in the current experiment to determine variable importance (including country context and algorithm choice). However, the result differed depending on which of the two random forest approaches was used (permutation versus node purity). Third, no conclusion is made as to the value and potential contribution of machine learning to prediction and targeting of social protection in the context of dynamic data, which may offer greater opportunity for improvements.

Annex 8A: Data

Data Annex

Country	Year	Observations	Constant USD	$PPP	Total (m)	0–14 (%)	Urban (%)	Density	Primary (%)	Unemployment	National	PPP 1.90	Gini
		Survey details	**GDP per capita**		**Population and education**					**Unemployment, poverty, and inequality**			
Comoros	2004	2,986	1,283	2,804	0.6	43	28	321	NA	5	36	14	56
Congo, Rep. of	2005	5,002	2,572	3,246	3.6	42	61	11	NA	20	44	53	47
Eswatini	2009	3,167	4,044	7,290	1.1	40	22	61	NA	27	49	42	52
Gambia, The	2010	4,670	861	2,347	1.8	45	56	177	39	9	35	25	44
Guinea-Bissau	2010	3,178	558	1,747	1.5	43	40	54	NA	3	62	67	51
Lesotho	2010	4,330	1,137	2,274	2.0	35	25	66	41	27	51	27	45
Liberia	2007	3,595	482	1,333	3.5	43	47	36	NA	2	57	69	37
Liberia	2014	4,101	586	1,621	4.4	42	49	45	NA	2	50	39	33
Mauritius	2012	6,705	8,580	18,015	1.3	21	41	619	67	7	5	1	39
Niger	2005	6,690	440	961	13.6	49	16	11	14	3	54	75	44
Niger	2007	4,000	446	973	14.7	50	16	12	14	2	50	72	37
Niger	2011	3,859	466	1,018	17.1	50	16	14	14	0	41	50	32
Niger	2014	3,627	516	1,126	19.2	50	16	15	14	1	37	45	34
Senegal	2011	5,953	1,263	2,707	13.0	44	44	68	22	10	36	38	40
South Sudan	2009	4,967	1,514	NA	9.1	44	18	NA	63	13	47	43	46
Togo	2011	5,491	553	1,269	6.6	43	38	121	30	2	48	54	46
Togo	2015	2,335	631	1,447	7.3	42	40	135	30	2	46	50	43

Source: World Development Indicators (WDI).
Note: $PPP line poverty line is $1.25 (PPP) or $1.90 a day (2011 PPP). GDP = gross domestic product; m = millions; NA = not available; PPP = purchasing power parity; USD = US dollars.

Annex 8B: Statistical Methodology and Supplementary Analysis

This annex includes the following:

1. An overview of the statistical methodology used to conduct the mixed model metadata analysis on the full set of simulations
2. Supplementary analysis supporting some of the results identified in the main text.

Mixed Model Methodology

Although it is still relatively easy to visualize the results of a single model and a country data set for a single performance measure, this rapidly becomes impractical once the analysis includes more algorithms, more country data sets, and more performance measures. Using a statistical model for a meta-analysis can more easily summarize a large number of results. This subsection describes the methodology for the benchmark experiment following Eugster, Hothorn, and Leisch (2012) and Hothorn et al. (2005).

Much like meta-analysis in research that combines results from multiple experiments to improve the estimates of effects and reduce uncertainty, the B resampling results of each performance measure p, algorithm a, and data set d consist of a single experiment. The results can then be pooled to gain a better estimate of the average algorithm performance or algorithm effect.

Modeling the resampling results with a mixed model allows for making statistical comparisons between algorithms and determining whether their differences in performance are statistically significant, while controlling for the fact there are B repeated measures for data sets and algorithms. As such, the observations are not independent and identically distributed and standard statistical approaches such as least squares regression are not valid. Ignoring this dependence between observations leads to increased type I errors (that is, false positives) and overly confident results (for example, too small standard errors).

For each performance metric p, the $D \cdot A \cdot B$ observations are modeled using a mixed-effects model. It is written as:

$$p_{dab} = \alpha_a + \beta_d + \beta_{da} + \varepsilon_{dab}$$

where:

$p=1,...,P$	performance measures
$d=1,...,D$	data sets
$a=1,...,A$	algorithms
$b=1,...,B$	validation samples.

The dependent variable is the performance metric p. On the right-hand side, α_a represents the candidate algorithms' mean performances, β_d are the mean performances on the data sets, β_{da} is the interaction of the data sets and algorithms, β_{db} is the effect of subsampling within data sets, and ϵ_{dab} is the systematic error.

Variable α_a is modeled as a fixed effect, while β_d is a random effect, as are β_{db} and ϵ_{dab}. The random effects follow $\beta_d \sim N(0, \ \sigma_1^2)$, $\beta_{db} \sim N(0, \sigma_2^2)$, and

$\epsilon_{dab} \sim N(0, \sigma_3^2)$, and the analysis can rely on the asymptotic normal and large sample theory (see Eugster, Hothorn, and Leisch 2012).
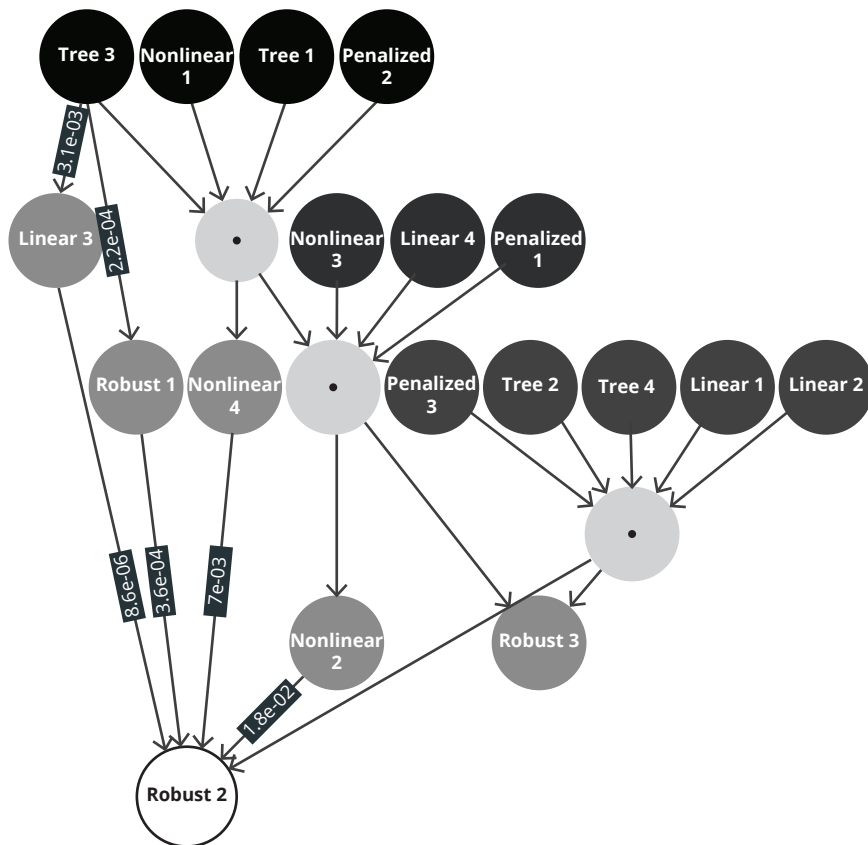
The model allows the following interpretation, of course, conditional on the domain D, for an algorithm a and a data set $d : \hat{\alpha}_a$ is the algorithm's mean performance, $\hat{\beta}_d$ is the data set's mean complexity, and $\hat{\beta}_{db}$ is the algorithm's mean performance difference from its mean performance conditional on the data set, or as Eugster, Hothorn, and Leisch (2012) put it, "how the algorithm likes the data set."

Supplementary Results

A test can be employed to determine whether any two algorithms' normalized root mean square errors (NRMSEs) are statistically different from each other.[27] There are different ways of thinking about algorithm performance: (1) which algorithms are better or worse than each of the others (pairwise statistical wins and losses), and (2) which algorithms are statistically better or worse when compared with all others. The Hasse diagram in Figure 8B.1 shows both of these. It is read from bottom (smallest values) to top (highest values). In this case, rather than maximizing a machine learning algorithm as is usually the case, the analysis is minimizing the NRMSE, so the best performers are at the bottom and an algorithm that dominates another has an arrow pointing to it from the one it dominates. Where no arrow (or series of arrows) connects two algorithms, they are not statistically different in performance (algorithms in the same statistical tier have the same color). While Tree 3 has the most net pairwise wins (seven arrows cumulatively away from it), it is not statistically different from 10 other algorithms (the black tier). The figure also highlights that on this performance measure, NRMSE relating to estimation of the entire distribution, robust models perform systematically worse as a category. These two different ways of thinking about performance—(1) which algorithms are better or worse than each of the others (pairwise statistical wins and losses), and (2) which algorithms are statistically better or worse compared with all the others (arrow tracing in the Hasse diagram)—are summarized in table 8B.1. The leftmost column shows the net number of pairwise statistical wins and losses each algorithm has. For example, Robust 1 has one win over Robust 3 but two losses to Tree 3 and Nonlinear 1 for net statistically significant wins of −1. However, the number of relevant wins is also presented, which is defined here as having a 5 percent lower NRMSE as a threshold for being meaningfully different. Under this definition, there are almost no net meaningful wins or losses for any algorithm. The right-hand side shows the equivalence classes for the algorithms; letters grouped under the same letter (where *a* is better than *b*) are not statistically different from each other (no arrow tracing). The table presents the results of the Hasse diagram in an

easily summarized, compact letter display, emphasizing that although Tree 3 has the most net wins, it is statistically not different from 10 other algorithms. The table also highlights that on this performance measure, NRMSE relating to estimation of the entire distribution, robust models perform systematically worse as a category.

Figure 8B.1   NRMSE Performance Pairwise Comparisons, by Algorithm, Using the Two-Sided Tukey Test

Source: Areias et al., forthcoming.
Note: Linear 1 = OLS without variable selection; Linear 2 = OLS with stepwise selection and Akaike Information Criterion; Linear 3 = OLS with stepwise selection and p-value criterion; Linear 4 = partial least squares; Linear 5 = logistic Lasso with variable selection; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression; Robust 3 = quantile regression with cross-validated quantile; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Nonlinear 4 = neural network; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees.

Table 8B.1 Compact Letter Display of Equivalence Classes for NRMSE Performance

Algorithm	Significant wins	Relative wins	Equivalence Class					
			a	b	c	d	e	f
Linear 2	4	0	a	b	c			
Linear 3	1	0		b		d	e	
Linear 4	4	0	a	b	c			
Robust 1	1	0		b		d	e	
Robust 2	−11	0						f
Robust 3	−10	0				d		f
Penalized 2	5	0	a	b				
Penalized 1	4	0	a	b	c			
Penalized 3	4	0	a	b	c			
Nonlinear 1	5	0	a	b				
Nonlinear 3	4	0	a	b	c			
Nonlinear 2	–4	0			c	d		f
Tree 1	5	0	a	b				
Tree 2	3	0	a			d		
Tree 3	7	2	a					
Tree 4	4	0	a	b	c			

Source: Areias et al., forthcoming.

Note: Linear 2 = OLS with stepwise selection and AIC; Linear 3 = OLS with stepwise selection and *p*-value criterion; Linear 4 = partial least squares; Robust 1 = robust linear regression with principal components; Robust 2 = quantile regression with stepwise AIC variable section; Robust 3 = quantile regression with cross-validated quantile and stepwise AIC variable section; Penalized 1 = Ridge regression; Penalized 2 = Lasso regression; Penalized 3 = elastic net; Nonlinear 1 = multivariate adaptive regression splines; Nonlinear 2 = *k*-nearest neighbors; Nonlinear 3 = support vector machines; Tree 1 = random forest; Tree 2 = random forest with quantile loss; Tree 3 = gradient boosted regression trees; Tree 4 = gradient boosted quantile regression trees. AIC = Akaike Information Criterion; OLS = ordinary least squares.

Table 8B.2 Summary of the Models and Their Characteristics

Model	Allows $n < p$	Pre-processing	Interpretable	Automatic feature selection	# Tuning parameters	Robust to predictor noise	Computation time
Linear regression†	✘	CS, NZV, Corr	✓	✘	0	✘	✓
Partial least squares	✓	CS	✓	O	1	✘	✓
Ridge regression	✘	CS, NZV	✓	✘	1	✘	✓
Elastic net/Lasso	✘	CS, NZV	✓	✓	1–2	✘	✓
Neural networks	✓	CS, NZV, Corr	✘	✘	2	✘	✘
Support vector machines	✓	CS	✘	✘	1–3	✘	✘
MARS/FDA	✓		o	✓	1–2	o	o
k-nearest neighbors	✓	CS, NZV	✓	✘	1	o	✓
Single trees	✓		o	✓	1	✓	✓
Model trees/rules†	✓		o	✓	1–2	✓	✓
Bagged trees	✓		✘	✓	0	✓	o
Random forest	✓		✘	o	0–1	✓	✘
Boosted trees	✓		✘	✓	3	✓	✘
Cubist†	✓		✘	o	2	✓	✘
Logistic regression*	✘	CS, NZV, Corr	✓	✘	0	✘	✓
Nearest shrunken centroids*	✓	NZV	o	✓	1	✘	✓
Naïve Bayes*	✓	NZV	✘	✘	0–1	o	o
C5.0*	✓		o	✓	0–3	✓	✘

Source: Kuhn and Johnson 2018.

Note: †regression only; *classification only; Symbols represent affirmative (✓), negative (✘), and somewhere in between (o); Corr = removing highly correlated predictors; CS = center and scaling predictors; MARS/FDA = multivariate adaptive regression splines/flexible discriminant analysis; NZV = removing near-zero predictors.

Notes

1. This is also called out-of-sample validation of in-sample models (referred to as cross-validation in the machine learning literature).
2. See Areias et al. (forthcoming) for discussion of a special case of *k*-fold validation called *leave-one-out cross validation*, where *k* is set to *n*, the number of observations in the data.
3. Thus, the assessment includes (1) linear regression with no variable selection technique (all variables included), (2) linear regression with stepwise variable selection based on the *p*-value criterion, (3) linear regression with stepwise variable selection based on the Akaike Information Criterion, (4) partial least squares, and (5) logistic regression with Lasso variable selection. The first three models listed are ordinary least squares (OLS) and vary only in terms of their approach to variable selection. Some PMT models do not filter variables to model (or do so as part of an ad hoc or expert judgment process outside the modeling). In this case, all the variables enter the OLS model and all the coefficients are retained. Alternatively, the second OLS model uses the stepwise approach discussed earlier to select variables and filters variables based on their *p*-values of contributing to the model. The third OLS model also uses stepwise, but it uses the Akaike Information Criterion.
4. The five robust models are (1) robust linear regression with principal components, (2) quantile regression, (3) quantile regression with cross-validated quantile, (4) quantile regression with Akaike Information Criterion variable selection, and (5) quantile regression with Akaike Information Criterion variable selection and cross-validated quantile.
5. The three penalized regressions are (1) Ridge regression, (2) Lasso regression, and (3) elastic net, which is a combination of Ridge and Lasso.
6. The four nonlinear models are (1) multivariate adaptive regression splines (MARS), (2) support vector machines (SVM), (3) *k*-nearest neighbors (KNN), and (4) neural networks.
7. Tree-based models are popular for three main reasons: (1) they generate conditions that are highly interpretable and easy to implement; (2) they do not require specifying the relationship between the explanatory and outcome variables ahead of modeling; and (3) they handle missing data and implicitly conduct variable selection. At the same time, two well-known weaknesses are (1) model instability, and (2) predictive performance that can be beaten by other approaches. Ensemble methods have been developed to address these issues. This chapter looks at both basic regression trees and ensemble approaches that build on them. Four variants of tree-based models are considered, which are ensemble models designed to take advantage of the tree-based approach while mitigating its weaknesses: (1) random forest, (2) random forest with quantile loss, (3) gradient boosted regression trees, and (4) gradient boosted quantile regression trees.
8. A resampling strategy (see Areias et al. [forthcoming]) was used to (1) draw 100 survey-weighted bootstrap samples of each data set (same samples used for all algorithms), (2) fit the algorithm to bootstrap samples and compute performance measures of the model using out-of-sample observations, and

(3) save 100 out-of-sample error estimates from the optimal parameter set for use in mixed-model analysis.

9. The countries and years used are the Comoros (2004), the Republic of Congo (2005), Eswatini (2009), The Gambia (2010), Guinea-Bissau (2010), Lesotho (2010), Liberia (2007, 2014), Mauritius (2012), Niger (2005, 2007, 2011, 2014), Senegal (2011), South Sudan (2009), and Togo (2011, 2015).

10. Earlier versions of the paper also examined quantile regressions with no variable selection and quantile regressions with cross-validated quantiles. They are omitted here for ease of exposition and their results do not change the main narrative.

11. A hybrid approach of the two can also be taken. For example, in Indonesia, the quota approach is used, but a maximum eligible PMT score threshold is also set to stop obviously nonpoor households from being included, even when the quota has not yet been met.

12. Root mean square error normalized to allow comparison across data sets with different standard deviations.

13. Inclusion and exclusion errors are not scale-invariant (similarly, neither are F_1 or F_2). For example, consider program A, which aims to cover the poorest 10 percent of the population, and program B, which aims to cover the poorest 30 percent. If targeting is random for both programs, then 1 in 10 of the poorest 10 percent will be covered under program A, meaning a 90 percent exclusion error. At the same time, 3 in 10 of the poorest 30 percent will be covered under program B, meaning a 70 percent exclusion error. Thus, the same targeting performance results in different exclusion (and inclusion) errors for programs of different coverage size. MCC is calculated to be independent of program size. For further discussion, see chapter 7, Powers (2007), and Wai-Poi (2011).

14. An upcoming revision of Areias et al. (forthcoming) will include full distributional characteristic results.

15. It also controls for the fact that there are 100 repeated measures for the data sets and algorithms and, as such, the observations are not independent and identically distributed, meaning that standard statistical approaches such as least squares regressions are not valid. Ignoring the dependence between observations leads to increased type I errors (that is, false positives) and overly confident results (that is, standard errors that are too small).

16. Robust models do not aim to minimize mean squared error but rather mean absolute error; as such, it is not surprising that they perform worse.

17. Hasse diagrams and compact letter displays are available for all performance measures in the original paper.

18. For NRMSE, a lower score is better (a smaller error), but for ease of reading, the NRMSE results have been reversed so that they can be read consistently with the other measures.

19. The results presented are the coefficients for each mode from the mixed model metadata analysis described in annex 8B. The OLS algorithm includes a preprocessing step that filters the dummified predictors so that there are no absolute pairwise correlations above 0.9 and also removes near zero variance predictors.

20. The results are similar for F_1. See annex 8B for mixed-method coefficient analysis.
21. Logistic regression with Lasso variable selection.
22. In an earlier version of the paper, when *quantregAIC* was included, it had contrasting performance on the F_2 and MCC measures.
23. Moreover, if different programs target significantly different sized populations, a single model that captures the entire distribution may be better than a single model that targets a specific eligibility threshold but that is then used for multiple programs at higher or different eligibility thresholds.
24. An annex in Areias et al. (forthcoming) presents the tabulated net statistical and relevant wins.
25. Technically, the model used in this chapter is a logistic model with Lasso variable selection, so it allows $n < p$ and is more robust for predictor noise since it uses penalization for variable selection, and it needs parameter tuning on both lambda (the penalty) and c (the threshold cutoff).
26. Kuhn and Johnson (2018, table A.1) also indicate automatic feature selection as a benefit of boosted models over logistic models. Here, the logistic is implemented with a simple Lasso variable selection, so feature selection does not seem to be an advantage in practice for boosted models.
27. A two-sided Tukey test.

References

Areias, Ana, Niall Keleher, Laura Ralston, and Matthew Wai-Poi. Forthcoming. "Big Data, Machine Learning and Prediction in the Targeting of Social Protection Programs." World Bank, Washington, DC.

Athey, Susan. 2018. "The Impact of Machine Learning on Economics." NBER Chapters. National Bureau of Economic Research, Cambridge, MA.

Eugster, Manuel J. A., Torsten Hothorn, and Friedrich Leisch. 2012. "Domain-Based Benchmark Experiments: Exploratory and Inferential Analysis." *Austrian Journal of Statistics* 41 (1): 5–26.

Hothorn, Torsten, Friedrich Leisch, Achim Zeileis, and Kurt Hornik. 2005. "The Design and Analysis of Benchmark Experiments." *Journal of Computational and Graphical Statistics* 14 (3): 675–99.

James, Gareth, Daniela Witten, Trevor Hastie, and Rob Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.

Kuhn, Max, and Kjell Johnson. 2018. *Applied Predictive Modeling*. New York: Springer.

Noriega-Campero, Alejandro, Bernardo Garcia-Bulle, Luis Fernando Cantu, Michiel A. Bakker, Luis Tejerina, and Alex Pentland. 2020. "Algorithmic Targeting of Social Policies: Fairness, Accuracy, and Distributed Governance." In Conference on Fairness, Accountability, and Transparency (FAT* '20), January 27–30, Barcelona, Spain. ACM, New York. https://doi.org/10.1145/3351095.3375784.

Powers, David M. W. 2007. "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation." Technical Report SIE-07-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia.

Russom, Philip. 2018. "2018 TDWI Checklist Report: The Automation and Optimization of Advanced Analytics Based on Machine Learning." Transforming Data With Intelligence, Renton, WA. www.tdwi.org.

Wai-Poi, Matthew. 2011. "Advanced Targeting Metrics." Background paper for World Bank (2012) *Targeting the Poor and Vulnerable in Indonesia*, World Bank, Washington, DC.

Wolpert, David H. 1996. "The Lack of *a priori* Distinctions between Learning Algorithms." *Neural Computation* 8 (7): 1341–90.