Are Vaccination Campaigns Misinformed?

Experimental Evidence from COVID-19 in Low- and Middle-Income Countries

Yannick Markhof
Philip Wollburg
Alberto Zezza

Abstract

Routine immunization coverage estimated in surveys often substantially differs from figures reported in administrative records, presenting a dilemma for researchers and policy makers. Using high-frequency phone surveys and administrative records from government sources in 36 low- and middle-income countries, this paper shows that such misalignment has also been common in the case of COVID-19. Across the sample, survey estimates exceed administrative figures by 47 percent on average, at times suggesting markedly different policy conclusions depending on the data source consulted. This pattern is particularly stark and consistent in Sub-Saharan Africa. To investigate the sources of this discrepancy, the paper presents results from six methodological experiments that vary survey design choices and documents their effect on estimated COVID-19 vaccine coverage. The results show that design choices matter, in particular the selection of respondents to be interviewed. However, phone survey estimates prove remarkably robust to several commonly claimed biases. After accounting for observed errors of representation and measurement in the survey data, there remains a nonnegligible, unexplained residual gap with administrative records. The paper provides indicative evidence of flaws and weaknesses in administrative data recording and reporting that affect reported vaccination rates and could contribute to this gap. The findings matter for past research on COVID-19 vaccination, future immunization efforts, and the design of robust data production systems on health topics.

Are Vaccination Campaigns Misinformed? Experimental Evidence from COVID-19 in Low- and Middle-Income Countries

Yannick Markhof [*][†][‡], Philip Wollburg [‡], Alberto Zezza [‡]

JEL codes: I14, I18, C81, C82, C83

Keywords: COVID-19, vaccination, survey data, administrative data

  * Corresponding author (markhof@merit.unu.edu)
† United Nations University, UNU-MERIT
‡World Bank, Development Economics Data Group

1. Introduction

Investments in large-scale vaccination efforts have led to drastic reductions in child mortality over the last decades (WHO 2020a). Much of this investment has targeted low- and middle-income countries, which will receive USD 1.2 billion in support between 2021 and 2025 from Gavi, a multi-donor initiative aiming to cut the share of children without any routine immunization in half by 2030 (Gavi, The Vaccine Alliance 2022b; 2021). Achieving further reductions in the share of unvaccinated children requires reliable data on vaccine coverage, which children remain unvaccinated, where and why (Galles et al. 2021; Danovaro-Holliday et al. 2021; Scobie et al. 2020; WHO 2020a; Cutts et al. 2016). A substantial share of funding is thus directed toward reliable and fit-for-purpose data systems (Gavi, The Vaccine Alliance 2022c; WHO 2020a).

A strong emphasis on data quality is a core principle of the WHO's Immunization Agenda 2030 and reflects the doubtful reliability of vaccine coverage estimates to date (WHO 2020a; Galles et al. 2021; Cutts et al. 2016). Specifically, it is not uncommon to see estimates of vaccine coverage differ by double digits between administrative and survey-based sources (Galles et al. 2021; Sandefur and Glassman 2015; Miles et al. 2013; Burton et al. 2009; Dykstra et al. 2019). In 2017, less than half of Gavi-supported countries reported survey estimates of vaccine coverage that were within 10 percentage points of administrative figures (Gavi, The Vaccine Alliance 2022c). Such misalignment is taken by donors as the core metric by which progress in data quality is evaluated (Gavi, The Vaccine Alliance 2022c) and is the subject of this study.

Previous research on this topic has focused on routine immunization and in-person data collection, comparing different data sources such as self-reports, home-based records, and health facility or administrative data (Dansereau et al. 2020; Sandefur and Glassman 2015; Miles et al. 2013; Lim et al. 2008). In these studies, the scope of experimentation to explore different design choices and their effect on the reliability of vaccine uptake estimates has been small. Differences in design choices can have a direct impact on the quality of data collected and affect the policy conclusions and investment recommendations drawn (De Weerdt et al. 2020a). This makes their assessment an important empirical question to investigate (Dillon et al. 2020).

In this study, we document that substantial misalignment between survey and administrative data is also pervasive in COVID-19 vaccine coverage figures. In a sample of 36 LMICs, survey-based estimates suggest COVID-19 vaccine coverage that is on average 47% higher than what is documented in administrative sources. This pattern displays distinct regional variation that is particularly striking and consistent in Sub-Saharan Africa.

We then systematically investigate the sources of this misalignment in Sub-Saharan Africa. Our empirical strategy exploits the unprecedented availability of high-frequency administrative data and concurrently collected (phone) survey data on COVID-19 vaccinations, which allows us to directly compare vaccine coverage estimates from these two sources at multiple points in time and in multiple countries. We conduct a series of (randomized) survey experiments using cross-country comparable, longitudinal (phone) surveys across five LMICs in Sub-Saharan Africa. Each experiment exogenously varies one aspect of the survey design that we hypothesize could give rise to the observed discrepancy between survey and administrative data-based estimates of the

COVID-19 vaccination rate. We broadly classify these design choices into questions of measurement and representation (Groves 1989; Groves and Lyberg 2010). Together, the design choices we study span some of the most common sources of error covered under the total (survey) error framework that has been the guiding paradigm for research on data quality (Amaya et al. 2020; Groves and Lyberg 2010; Groves 1989).

We find that the substantial misalignment between survey and administrative data-based estimates of vaccine coverage is only partly explained by measurement and representation issues in survey data. After accounting for observed errors of representation and measurement at the household and individual level, there remains a statistically significant gap of 6%-52% between survey and administrative estimates in our study countries. Our findings indicate that design choices determine the reliability and suitability of the survey data for capturing vaccine coverage. At the same time, our explorative analysis suggests administrative data also suffers from flaws and inaccuracies. Without giving due consideration to both survey design and potential inaccuracies in administrative records, both data sources can lead to different findings and different policy conclusions. Our findings have implications for research and policy making on COVID-19 vaccination but also for future data collection on routine immunization and supplementary immunization activities.

Our study relates and contributes to several strands of literature. Conceptually, we frame our analysis as part of the literature on total survey error and total survey quality (Amaya et al. 2020; Groves and Lyberg 2010; Biemer 2010) as well as studies analyzing the role of design choices in producing credible research and value for development (De Weerdt et al. 2020a; Jolliffe et al. 2023). This research emphasizes the sensitivity of both survey and administrative data to non-classical sources of measurement error. In this framing, conducting robust research is akin to an optimization problem in which the researcher tries to maximize credible knowledge subject to budget and data quality constraints (Dillon et al. 2020). Improvements in data quality provide opportunities to improve the quality of inference conducted. In line with Dillon et al. (2020), we argue that research on vaccination has paid insufficient attention to the sensitivity of data quality to design choices. Our study addresses this gap.

Secondly, for vaccination data specifically, we contribute to a body of research documenting and analyzing misalignment between survey-based and administrative sources (Wolter et al. 2022; Nguyen et al. 2021; Galles et al. 2021; Bradley et al. 2021; Sandefur and Glassman 2015; Miles et al. 2013; Murray et al. 2003; Lim et al. 2008; Burton et al. 2009). This literature stresses that both data sources are subject to potential errors. The strength of administrative data is its spatial granularity and frequency, but it often suffers from numerator (number of people vaccinated) and denominator (size of the target population) issues.[1] Survey data, for example from large household survey programs such as the Demographic and Health Surveys (DHS), Multiple Indicators Cluster Survey (MICS), or Living Standards Measurement Study (LSMS) is independent of public record keeping and can also capture vaccination obtained through private or non-governmental providers

---

[1] For example, numerator issues may occur when administrative data misses out on vaccinations conducted through the private health care sector and or when aggregating figures from lower administrative levels. Denominator issues may relate to outdated census data or inaccurate population projections (Burton et al. 2009).

(which administrative records may miss). Further, surveys can provide estimates when the size of the target population is unknown and present the opportunity to collect rich additional data. However, survey data is typically collected in less frequent intervals, representative at coarser administrative levels, and can be subject to measurement error (Danovaro-Holliday et al. 2021; Althubaiti 2016; Burton et al. 2009). In the absence of conclusive evidence in favor of either data source, the official WHO/UNICEF estimates of vaccine coverage have used an arbitration procedure whereby administrative data is used if the discrepancy with survey data is smaller than 10 percentage points and survey data otherwise as long as it is deemed "credible" (Burton et al. 2009; 2012; Brown et al. 2013). Other research and policy has preferred survey over administrative data for its purported independence and higher accuracy (Dykstra et al. 2019; Cutts et al. 2016; Sandefur and Glassman 2015; Lim et al. 2008; Gavi, The Vaccine Alliance 2022a) and in rarer cases (typically when researchers had some control over the quality of administrative records) used only administrative data or both data sources complementarily (Banerjee et al. 2010; Barham and Maluccio 2009; Banerjee et al. 2019).

Evidence on the potential size and direction of the misalignment between survey and administrative is scarce in the context of COVID-19 and focuses on the United States or Germany (Wolter et al. 2022; Bradley et al. 2021; Nguyen et al. 2021). These studies found survey estimates to exceed administrative data with some evidence for errors of representation and measurement in the survey data. To the best of our knowledge, ours is the first large scale study investigating this issue in the context of LMICs and testing a comprehensive range of survey design choices.

A third stream of literature is applied (typically microeconomic) research using survey and sometimes administrative vaccination data in LMICs to inform policy. In these studies, vaccine uptake is usually the outcome of interest, regressed on some hypothesized determinant of vaccination behavior or policy intervention. Examples in development economics abound but can be found, for instance, in the literature on cash transfer interventions (Haushofer and Shapiro 2016; Chandir et al. 2022; Kusuma et al. 2017; Celhay et al. 2021; Barham and Maluccio 2009; De and Timilsina 2020; Benedetti et al. 2016; Debnath 2021), the literature on improving the quality and utilization of health services (Banerjee et al. 2010; Björkman and Svensson 2009; Christensen et al. 2021; Blimpo et al. 2022) and many others (Cockx 2022; Levine et al. 2021; Aggarwal 2021; Keats 2018; Palloni 2017; Adhvaryu et al. 2019; Stoop et al. 2019; Miller and Urdinola 2010; Banerjee et al. 2019).[2] Most directly, our study relates to a growing body of research that specifically studies COVID-19 vaccination and aims to inform vaccination campaigns with estimates on vaccine acceptance (Lazarus et al. 2021; 2022; 2023; Solís Arce et al. 2021; Kanyanda et al. 2021; Wollburg et al. 2023; Dayton et al. 2022) and uptake (Wollburg, Markhof, et al. 2022; H. M. Reza et al. 2022). Our study informs this research by testing the reliability of data underlying it.

A final stream of literature is methodological research on phone surveys in LMICs. This research has seen a significant increase in interest since the COVID-19 pandemic and led to calls for the integration of phone surveys into routine data collection schedules in health, economic,

---

[2] See also the literature reviews of Andreas et al. 2022, Jain et al. 2022, Machado et al. 2021, Kaufman et al. 2018, Jacobson Vann et al. 2018, Oyo-Ita et al. 2016, Harvey et al. 2015, Groom et al. 2015, and Saeterdal et al 2014.

agricultural research, and beyond (Gourlay et al. 2021; Zezza et al. 2022; Glazerman et al. 2023). In this regard, phone surveys, such as those we study, address two common criticisms of the usefulness of (in-person) survey data for health policy: their low temporal frequency and sparse coverage in conflict-affected or hard-to-access areas. Our study explores the reliability of phone survey data for vaccine research on COVID-19 and beyond.

The remainder of this paper proceeds as follows. Section 2 describes our data and Section 3 our empirical strategy. Section 4 presents our results. Section 5 discusses the implications of our results for health policy and research. Section 6 concludes.

2 Data

We use data from three sources: phone surveys, in-person surveys, and administrative records.

Phone surveys have become key tools to fill information gaps when in-person data collection came to a near complete halt during the COVID-19 pandemic (Wollburg, Contreras, et al. 2022). As a result, they have become widespread and enabled repeated experimentation at high frequency for the purpose of this analysis (Gourlay et al. 2021; Glazerman et al. 2020).

Specifically, we use data from longitudinal and cross-country comparable national phone surveys implemented between March 2021 and January 2023. These multi-topic phone surveys were conceived in order to track the effects of the COVID-19 pandemic in the absence of in-person data collection (Himelein et al. 2020). Our experimentation with survey design choices draws on five of these surveys in Sub-Saharan Africa that were supported by the Living Standards Measurement Study (LSMS) team at the World Bank and implemented by the respective National Statistical Offices. These surveys are re-contact surveys, drawing their samples from the latest nationally representative, in-person LSMS-ISA household survey conducted in each country before the pandemic. As part of the LSMS-ISA surveys, phone contact numbers of all household members (where available) as well as from a reference contact such as a neighbor were collected (Gourlay et al. 2021). The list of households with a phone contact, or a random subset of it, constituted the sample to be contacted for the phone surveys and covered between 73% (Malawi) and 99% (Nigeria) of households included in the in-person LSMS-ISA survey. This approach also led to response rates that compare favorably to other phone surveys, especially those employing random digit dialing (Dillon et al. 2021; Gourlay et al. 2021; Henderson and Rosenbaum 2020). It also allowed to draw on a rich set of household characteristics to attenuate coverage biases through re-weighting techniques (Ambel et al. 2021; Himelein et al. 2020; Brubaker et al. 2021). The respondent for each phone survey interview was purposively selected as an adult (15+) household member that is knowledgeable of the affairs of the household, typically the household head.

The data we use in this study comes from a harmonized survey module on COVID-19 vaccination that was first fielded in August 2020 and then periodically repeated. The content of the survey module on vaccines varied over time in response to changing data demands as vaccination campaigns progressed. The focus of this paper is a question on whether the respondent had been vaccinated for COVID-19 which was included in the survey module after COVID-19 vaccines

became available. Our total sample comprises of 57 rounds of data across 36 countries amounting to over 94,000 individual-level data points (Table A1). Our experimental results focus on five of these 36 countries that are located in Sub-Saharan Africa.

The second source of data we draw on is a short survey on COVID-19 vaccination collected in-person as part of the Ethiopia Socioeconomic Survey (ESS 5), a nationally representative household survey that was implemented between April and June 2022 by the Ethiopia Statistical Service with support from the World Bank's LSMS program. This survey contained a similar module as the phone surveys and collected information on the vaccination status of all household members.

The source of administrative data for our study is the Our World in Data (OWID) COVID-19 vaccination dataset (Mathieu et al. 2021) that compiles administrative data on COVID-19 vaccine coverage. Amongst others, the dataset contains information on the number of total doses administered, the share of the country population that has received at least one dose, and the share of the population that is fully vaccinated.[3] It covers the period from December 2020 when the first COVID-19 vaccines achieved approval and is regularly updated as new data becomes available on a per-country basis. The data is compiled from country reports (such as government websites, dashboards, or the social media accounts of national authorities) and in some cases third-party aggregators (where national authorities do not publish data in a machine-readable format) and is regularly audited for inconsistencies and technical errors. The Our World in Data COVID-19 vaccination dataset has been extensively used during the pandemic, for example to supply the data for the WHO's official COVID-19 dashboard, by global media outlets, and for social science and epidemiological research (Mathieu et al. 2021). The dataset is publicly available through the Our World in Data GitHub repository.[4]

We additionally access a second source of administrative data stemming from the WHO's COVID-19 vaccination dashboard (WHO 2020b). The dashboard does not provide longitudinal information for public access but reports the latest available COVID-19 vaccine coverage figures at the time of data access (April 2, 2023, in our case).

Lastly, we use data from the World Bank's Statistical Performance Indicators (SPI) available through the World Bank's Open Data library (World Bank n.d.). The SPI is a composite index between 0 – 100 scoring countries' statistical systems across the five pillars of data use, data services, data products, data sources, and data infrastructure (Dang et al. 2023). To capture the performance of administrative data systems in particular, we also use the SPI's indicator of administrative data capacity (Dimension 4.2) that records the availability of Civil Registration and Vital Statistics (CRVS).

---

[3] Defined as the share of the total population that has completed the initial vaccine protocol: two doses or more of a two-dose vaccine (such as Comirnaty by BioNTech and Pfizer) or one shot of a single-shot vaccine (such as Janssen by Johnson & Johnson).

[4] https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations

3 Empirical Strategy

Our analysis is based on comparing the national COVID-19 vaccine coverage rate[5] for a given population of interest estimated from survey data to the coverage rate reported in administrative data. Neither source can be regarded ex-ante as bias free or (close to) the "true" rate, so that we cannot observe data accuracy directly. Instead, we observe how different survey estimates vary in response to design choices and vis-à-vis the administrative data. This approach is common in the measurement literature in the absence of an objective truth against which estimates under different design choices can be benchmarked (Bardasi et al. 2011; Laajaj and Macours 2021; Das et al. 2012; Beaman and Dillon 2012; De Weerdt et al. 2020b). It is also in line with health policy practice in which the size of the gap between survey and administrative estimates of vaccine coverage is taken as the key metric by which data quality as a whole is judged (Gavi, The Vaccine Alliance 2022c).

An important caveat is that the two data sources usually refer to different reference populations. The phone surveys generally cover the adult population (aged 15+)[6] whereas the administrative data is reported for the entire country population. As age-disaggregated administrative data is not consistently available across our sample, we assume that the administrative data contains no vaccinated children (younger than 15 years). This assumption is likely more accurate at the start of the pandemic but will possibly underestimate the true gap between survey data and administrative records as countries lowered the age threshold for COVID-19 vaccination.[7]

We structure our analysis along six hypotheses and associated predictions that we empirically test by randomly varying one aspect of the survey design at a time (Table 1). We conduct each analysis on a per-country basis. In the following, we introduce each hypothesis, its predicted effect on the estimated COVID-19 vaccine coverage rate and our empirical strategy to test it. Our first two hypotheses relate to possible errors of *representation* whereas the remaining four hypotheses cover errors of *measurement*.

---

[5] Defined as the share of the total population that has received at least one dose of a COVID-19 vaccine.

[6] Some phone and in-person survey rounds collect information on the vaccination status of all household members, rather than just the survey respondent, in which case the survey and administrative data refer to the same reference population. We analyze these survey rounds below.

[7] Among the countries in which we conduct our survey experiments, children under the age of 16 started to be vaccinated in Ethiopia (from Nov 17, 2021), Malawi (from December 4, 2021), and Uganda (from July 26, 2021). Burkina Faso and Nigeria had not started vaccinating children yet at the end date of our data collection (Mathieu et al. 2021).

Table 1: Experimentation with Design Choices

Hypothesis	Prediction	Survey Data Source	Reference Admin Rate
Errors of representation			
H1: Household sample selection	P1: Households included in the phone survey will display higher rates of vaccine uptake compared to the general population in a nationally representative sample.	ESS 5 in-person vaccination survey in Ethiopia (April – June 2022)	GP: June 5, 2022
H2: Respondent selection	P2.1: Respondent selection at random will significantly reduce estimated vaccine uptake compared to purposive selection.	Phone survey rounds in Burkina Faso (Sep '22), Malawi (Sep '22) and Uganda (Aug '22)	AP: Oct 2, 2022 (BFA), Sep 4, 2022 (MWI, UGA)
	P2.2: Eliciting the vaccination status of all household members (via proxy-reporting) will significantly reduce estimated vaccine uptake compared to purposive selection.	as above	GP: same dates
	P2.3: Randomly selecting a household member to be interviewed or collecting (proxy-reported) data on all household members will lead to statistically indistinguishable estimates of vaccine uptake.	as above	n/a
Errors of measurement			
H3: Survey mode	P3: In-person surveying will display reported vaccine uptake similar to administrative records.	ESS 5 in-person vaccination survey in Ethiopia (Apr – Jun '22)	GP (all members of phone survey HHs) and AP (main respondents): June 5, 2022
H4: Panel conditioning	P4: Respondents interviewed for the first time, as opposed to repeat respondents, will display lower rates of vaccine uptake.	Phone survey round in Nigeria (Jan '22)	AP: Jan 16, 2022
H5: Proxy reporting	P5: Self-reported vaccine uptake will be significantly lower than proxy-reported vaccine uptake.	Phone survey rounds in Burkina Faso (Oct '22), Malawi (Sep '22) and Uganda (Aug '22)	AP: Oct 2, 2022 (BFA), Sep 4, 2022 (MWI, UGA)
H6: Experimenter demand	P6.1: Deliberately inducing positive experimenter demand will lead to significantly higher reported vaccine uptake. P6.2: Deliberately inducing negative experimenter demand will lead to significantly lower reported vaccine uptake.	Phone survey rounds in Burkina Faso (Jan '23) and Ethiopia (Jan '23)	AP: Feb 19, 2023 (BFA), Jan 22, 2023 (ETH)

Note: The table summarizes the empirical hypotheses we test in the vaccine survey data, the associated prediction(s) and survey data source. It also indicates the reference administrative coverage rate to which estimates can be compared. The date of the administrative rate refers to the admin data point closest to the time of survey completion. GP = General population, i.e. the entire population irrespective of age; AP = Adult population, i.e. the coverage rate calculated for the population aged 15 and above by assuming no vaccinations among younger persons.

3.1 Errors of representation

<u>Hypothesis 1: Household sample selection effects</u>

In the absence of universal phone ownership in our study countries, it is possible that phone survey samples overrepresent certain population groups such as better-off and urban households (Ambel et al. 2021; Brubaker et al. 2021). Similarly, non-response can lead to selective attrition from the sample, potentially affecting its representativeness for the general population. Even though our sampling strategy (collecting phone numbers from reference contacts) and re-weighting approach has been found to mitigate these issues, it is conceivable that some sample selection bias at the household level remains (Ambel et al. 2021; Gourlay et al. 2021). If selection into our sample (either through coverage bias in the list of households with phone numbers or through non-response) is correlated with vaccine uptake, estimated coverage will be biased. We thus formulate the first hypothesis for the divergence between survey and administrative vaccine coverage estimates.

Hypothesis 1: Phone surveys overrepresent population groups that are more likely to be vaccinated.

The COVID-19 vaccination module implemented as part of the Ethiopia Socioeconomic Survey (ESS 5) gives us the opportunity to test this hypothesis. This module collected information on vaccine uptake for a nationally representative sample of households of which our phone survey sample is a subset. While sample selection effects may affect the phone survey sample, they should be mostly absent in a fully nationally representative sample. Therefore, we can compare estimated vaccine uptake among the sample of phone survey households (interviewed during the ESS 5) to estimated vaccine uptake within the whole, general population sample contained in the ESS 5. We formulate the following prediction.

Prediction 1: Households included in the phone survey will display higher rates of vaccine uptake compared to the general population in a nationally representative sample.

Empirically, we run the following OLS regression.

$$V_{is} = \alpha + \beta D_{is} + \varepsilon \qquad (1)$$

Where V_{is} is a dummy for whether individual i part of sample s has been vaccinated with at least one jab of a COVID-19 vaccine. Subscript s denotes the sample to which individual i belongs. All individuals are part of the general population sample whereas only some are living in households sampled for the phone survey. Some individuals will thus appear twice in the data, once in the phone sample and once in the general population sample. Our estimate of interest is β, the coefficient on $D_{i,s}$ which is a dummy variable for whether observation i,s is part of the phone survey sample. If our prediction is confirmed, β should be statistically significant and positive.

Hypothesis 2: Respondent selection effects

Our second hypothesis relates to how respondents were chosen within phone survey households. The phone surveys typically interviewed one respondent per household per survey round who was selected to be knowledgeable across the different topics covered in the survey (the "main respondent"). Often this was the household head. As a result, respondent selection was purposive, not random, and overrepresented male, older, and more educated respondents relative to the general population (Brubaker et al. 2021). These traits, and other unobservables, may correlate with vaccine uptake and bias vaccine coverage estimates. We thus posit our second hypothesis.

Hypothesis 2: Purposive (non-random) respondent selection within the household overrepresents individuals that are more likely to be vaccinated.

We test this hypothesis in two ways. First, we randomly selected a respondent alongside the purposively selected ("main") respondent during one wave of the phone surveys in three countries (Burkina Faso, Malawi, Uganda). A respondent selected at random from among the members of the household is expected to be representative of the general population of eligible individuals, that is, all adults (15+) within households with access to a phone. Random selection of any eligible household member allows for the possibility that the purposively selected main respondent and the randomly selected respondent are the same individual. In this case, we conducted a single interview and count the observation both for the main respondent estimates and for the random respondent estimates.[8] Second, we asked the main respondent to report the vaccination status of all household members on their behalf ('proxy reporting'). This gives us the following information within each household: (i) the vaccination status of the purposively selected respondent; (ii) the vaccination status of a randomly selected household member (possibly, but not necessarily a different member than in (i); (iii) the vaccination status of all household members as reported by the purposively selected respondent. We formulate the following three predictions.

Prediction 2.1: Respondent selection at random will significantly reduce estimated vaccine uptake compared to purposive selection.

Prediction 2.2: Eliciting the vaccination status of all household members (via proxy-reporting) will significantly reduce estimated vaccine uptake compared to purposive selection.

Prediction 2.3: Randomly selecting a household member to be interviewed or collecting (proxy-reported) data on all household members will lead to statistically indistinguishable estimates of vaccine uptake.

To test these predictions, we run the following regressions.

$$V_{ir} = \alpha + \beta_1 D_{ir} + \varepsilon \tag{2}$$

$$V_{ir} = \alpha + \beta_2 T_{ir} + \varepsilon \tag{3}$$

---

[8] In these cases, the same individual will thus appear twice in the data, once as purposively selected respondent and once as a randomly selected respondent.

$$V_{ir} = \alpha + \beta_3 Z_{ir} + \varepsilon \qquad\qquad (4)$$

Where $V_{i,r}$ is again a dummy for whether respondent i selected by method r is vaccinated. The subscript r denotes whether V was obtained by interviewing a purposively selected respondent vs. a randomly selected respondent (equation 2); whether V was obtained from the purposively selected respondent vs. as part of eliciting the (proxy-reported) vaccination status of all household members aged 15 and above (equation 3)[9]; or whether V was obtained through random selection of a respondent vs. from the proxy reported vaccination status of all individuals in the household aged 15+ (equation 4). Our estimates of interest are β_1, β_2, and β_3, the coefficients on $D_{i,r}$, $T_{i,r}$, and $Z_{i,r}$, respectively, which are dummies for whether observation i,r is part of the randomly selected sample of respondents as opposed to the purposively selected sample ($D_{i,r}$ in equation 2); part of the sample for which a proxy reported vaccination status was collected as opposed to the sample of purposively selected respondents ($T_{i,r}$ in equation 3); or part of the randomly selected sample as opposed to the proxy-reported sample ($Z_{i,r}$ in equation 4); . If our predictions are confirmed, β_1 and β_2 should be significant and negative. Additionally, we would expect β_3 to be insignificant as the randomly selected respondent is sampled from the list of all household members aged 15 or older. To facilitate comparisons with the administrative data (without the need for assumptions regarding non-coverage of children), we additionally use the proxy reported vaccination status of all household members (i.e. not just those aged 15+) to estimate general population vaccine coverage.

3.2 Measurement errors

Hypothesis 3: Survey mode effects

Our third hypothesis relates to differences in survey modes. Concretely, it is conceivable that interviews conducted over the phone lead to different estimates than interviews conducted in-person. For example, respondents may be less inclined to answer truthfully when surveyed by an enumerator over the phone, pay less attention to the questions asked, and enumerators cannot ascertain the circumstances under which the interview takes place (e.g., other people present in the room, respondents getting distracted, etc.). This may lead to different reported vaccination rates in phone surveys than in in-person surveys. It is not obvious ex-ante whether phone or in-person surveys would find higher vaccination rates. However, for mode effects to explain (part of) the observed misalignment between surveys and administrative records, phone surveys would have to induce more respondents to report to be vaccinated. Our third hypothesis thus states the following.

Hypothesis 3: Asking respondents for their vaccination status over the phone as opposed to in person leads to greater reported vaccination rates.

To test this hypothesis, we again rely on the vaccination survey administered in person as part of the ESS 5 in Ethiopia. An ideal setup to test this hypothesis directly would be to randomly assign respondents either to the in-person survey or to the phone survey and conduct interviews around the same time. We do not have this kind of setup: we interview the same respondents both in

---

[9] This includes the purposively selected respondent's vaccination status which is a self- rather than a proxy-report.

person in the ESS 5 context and over the phone as part of the high frequency phone surveys, but there is a lag of nine months between these interviews, rendering the direct comparison of estimated vaccination rates difficult. Instead, we propose a more indirect test of this hypothesis through a comparison of vaccination rates according to administrative records to vaccination rates estimated in the in-person survey. We argue that if the phone survey mode effects were responsible for the observed misalignment between surveys and administrative records, the misalignment should disappear in in-person surveys. The prediction we test is the following.

Prediction 3: In-person surveying will display reported vaccine uptake close to administrative records.

Testing this prediction involves running a regression of vaccine uptake on a constant (to estimate the survey-based vaccine coverage rate) and performing a simple t-Test that tests the equality of the estimate obtained relative to the administrative coverage rate at the time of surveying.[10]

$$V_i = \alpha + \varepsilon \qquad (5)$$

$$t = \frac{\alpha - A}{\frac{s}{\sqrt{n}}} \qquad (6)$$

where V_i is the vaccination status of individual i, α is a constant, A is the COVID-19 vaccine coverage rate as reported in administrative data and t the test statistic obtained from a standard t-Test. If our prediction is confirmed, comparing t to the critical value of a Student t-distribution will no longer reveal a statistically significant difference between in-person survey estimates and the administrative data.

Hypothesis 4: Panel conditioning/Survey participation effects

The phone surveys we conduct are longitudinal, meaning that the same households (and often the same respondents) are interviewed multiple times regarding their COVID-19 vaccination status, willingness to get vaccinated, and related information. It is possible that this leads to a behavioral response in which respondents become more likely to get vaccinated (or report to be vaccinated) as they are repeatedly interviewed over time. Such 'panel conditioning' effects have been hypothesized to affect survey data across a wide range of topics but require experimental data for reliable identification (Struminskaya and Bosnjak 2021). Our hypothesis reads as follows.

Hypothesis 4: Repeatedly interviewing respondents on the topic of COVID-19 vaccination makes them more likely to get vaccinated or to report to have been vaccinated.

---

[10] As data collection for the in-person survey covered a period of multiple months, we conservatively take the administrative coverage rate from the date closest to the end of data collection for our survey. At this point in time, over 99% of data had been collected.

We test this hypothesis by exploiting a sample expansion of the phone survey in Nigeria after the first 12 rounds of data collection.[11] Initially, the sample of households selected for the phone survey had constituted of a randomly selected subset of all households for which a phone number was available from the latest in-person LSMS-ISA survey: 4,934 households had a phone contact, 3,000 of whom were randomly selected and 1,950 were successfully interviewed in rounds 1-12 of the phone survey from April 2020 to April 2021 (1,050 were not interviewed due to non-response and failed contact). Starting from round 13 of data collection in November 2021, the remaining 1,934 households with an available phone number were contacted as well to expand the sample. This setup provides us with a (randomly selected) sample of respondents who had been previously interviewed and a (randomly selected) sample of respondents who were interviewed for the first time in round 13, and we can compare their reported vaccination status. We make the following prediction.

Prediction 4: Respondents interviewed for the first time, as opposed to repeat respondents, will display lower rates of vaccine uptake.

To test our prediction, we run the following OLS regression.

$$V_i = \alpha + \beta D_i + \varepsilon \tag{5}$$

where $D_{i,s}$ is a dummy variable denoting whether individual i has been previously interviewed. In line with our prediction, we expect β to be significant and positive.

Hypothesis 5: Proxy reporting biases

When testing for respondent selection biases, one approach we explore is asking the purposively selected (main) respondent to report on the vaccination status of the remaining household members. However, such proxy reporting may be inaccurate (Davin et al. 2019; Li et al. 2015; Triplett 2010; Mosely and Wolinsky 1986). For errors in proxy reporting to drive the gap between survey estimates and administrative records, proxy reports would need to be systematically biased upwards, that is, main respondents would need to overstate vaccine uptake among the members of their household. We thus formulate the following hypothesis.

Hypothesis 5: Proxy reporting of other household members' vaccination status by the interviewed respondent overestimates true vaccine uptake.

We test this hypothesis by comparing two separate reports of the vaccination status for the same person: (i) as self-reported by the randomly selected respondent; and (ii) as proxy-reported by the purposively selected (main) respondent (see Hypothesis 2). This allows us to check the alignment of proxy reports with self-reports.[12] Our prediction is as follows.

---

[11] Out of these 12 rounds, three had collected data on COVID-19 vaccination.
[12] This excludes cases where the randomly selected respondent was identical to the purposively selected (main) respondent. In these cases, self-reported information equals "proxy-reported" information by definition.

Prediction 5: Self-reported vaccine uptake will be significantly lower than proxy-reported vaccine uptake.

To test this prediction, we run the following OLS regression on the pooled sample of self- and proxy-reported information for the same individuals.

$$V_{i,z} = \alpha + \beta D_{i,z} + \varepsilon \tag{5}$$

where z denotes whether the observation for individual i is a self- or proxy-report. $D_{i,z}$ is a dummy variable denoting whether vaccination status V was obtained via self-reporting. If our prediction is confirmed, we should observe β to have a negative sign and be significant. This would imply that main respondents' proxy reports overstate actual vaccine take-up.

Hypothesis 6: Experimenter demand effects

The final hypothesis we explore relates to the veracity of the information provided by respondents regarding their own vaccination status. Sensitive questions may induce respondents to not answer truthfully but in ways that they believe may protect their privacy, please the enumerator, or conform to what is "socially desirable". We hypothesize that asking about one's COVID-19 vaccination status may elicit such "experimenter demand" effects (de Quidt et al. 2018).[13]

Hypothesis 6: Respondents misreport their vaccination status (falsely claiming to be vaccinated) to conform with (perceived) socially desirable behavior or enumerator expectations.

In the absence of an objective way to verify the truthfulness of self-reports over the phone (e.g through home-based vaccination records[14]), we use a technique proposed by de Quidt et al. (2018) which we adapt to the context of COVID-19 vaccination and phone interviewing. The premise is to introduce random variation in the degree of experimenter demand into the sensitive survey question and to use this information to bound the potential bias stemming from experimenter demand in the survey-based estimate of vaccine take-up. Concretely, this involves making experimenter expectations explicit to random subsets of respondents by either telling respondents that the enumerator expects most to be vaccinated, or most to not be vaccinated.. Furthermore, part of the sample received the standard questionnaire design without additional information on enumerator expectations. Table 2 summarizes the introduction text read out to respondents in each treatment arm.

Table 2: Introduction text read out to different treatment groups

Treatment Group	Introduction text read out by the enumerator

---

[13] Another form of misreporting would arise from a situation in which respondents expect a tangible benefit from concealing their true vaccination status. We try to mitigate these issues in our survey by adding a disclaimer at the start of the vaccination module that answers will not be used to determine the respondent's eligibility status to receive a COVID-19 vaccine, nor to provide them with a COVID-19 vaccine.

[14] When asking our respondents whether they have proof for vaccination and if yes, of what type, the overwhelming majority report possessing a vaccination card. However, enumerators could not verify this information within the constraints of a phone interview.

T1 (Standard phrasing)	"Now, I'd like to ask you some questions to find out how many people have been vaccinated against COVID-19 in [country]. This will not determine your eligibility to receive a COVID-19 vaccine."
T2 (Positive demand)	Same as T1 with the addition: "[...] Many people in [country] have already been vaccinated against COVID-19. We thus expect that many of our survey respondents will tell us that they have been vaccinated."
T3 (Negative demand)	Same as T1 with the addition: "[...] Many people in [country] are not vaccinated against COVID-19. We thus expect that many of our survey respondents will tell us that they are not vaccinated."

We make the following predictions.

Prediction 6.1: Deliberately inducing positive experimenter demand will lead to significantly higher reported vaccine uptake.

Prediction 6.2: Deliberately inducing negative experimenter demand will lead to significantly lower reported vaccine uptake.

We then use estimated vaccine uptake to put bounds on the potential effect of experimenter demand on our estimate of interest. Concretely, the rate estimated from group T2 (for which experimenter demand to report being vaccinated was artificially high) should give an upper bound and the one estimated from group T3 (for which experimenter demand to report *not* being vaccinated was artificially high) a lower bound of true uptake in our sample. We can further compare this to estimated vaccine uptake among the group receiving the original survey question (T1) that was not framed in any particular light (and of which it is expected that the degree of (inadvertent) experimenter demand will lie between the bounds given by T2 and T3).

By introducing stronger positive (T2) and stronger negative (T3) reinforcement than is hypothesized to be present in the non-framed version of the question (i.e. the original questionnaire design, T3), it is assumed that the true, unbiased estimate of vaccine uptake lies in between the boundaries demarcated by vaccine uptake under T2 and T3.

To test our predictions, we can estimate the following regression.

$$V_i = \alpha + \beta_1 T_{2i} + \beta_2 T_{3i} + \varepsilon \tag{5}$$

where T_2 and T_2 are dummy variables for individual i belonging to treatment group 2 or 3, respectively. Our predictions imply that β_1 should be significant and positive whereas β_2 should be significant and negative.

Following the approach by de Quidt et al. (2018), we can further use the information from T2 and T3 to calculate a "demand robust" confidence interval on the standard, survey-based estimate of the vaccination rate (T1) that takes into account both uncertainty due to sampling error *and* the additional uncertainty of possible demand effects.

Provided that key identifying assumptions are met,[15] the size of the estimated demand-robust confidence interval on the vaccination rate specifies the range of values a survey-based estimate of vaccine take-up that is free from experimenter demand would fall into.

Finally, we can compare the point estimate of vaccine uptake under the standard questionnaire design (T1), along with its demand-robust confidence interval, to the vaccination rate as reported in administrative data sources. The larger the demand-robust confidence interval on our survey-based estimate of vaccine uptake, the larger the potential importance of enumerator demand in explaining misalignment between both data sources. Statistically significant differences between the administrative data and our survey-based estimate, with demand-robust confidence intervals, would suggest that enumerator demand effects alone cannot account for the difference between our survey-based estimates and the administrative data.
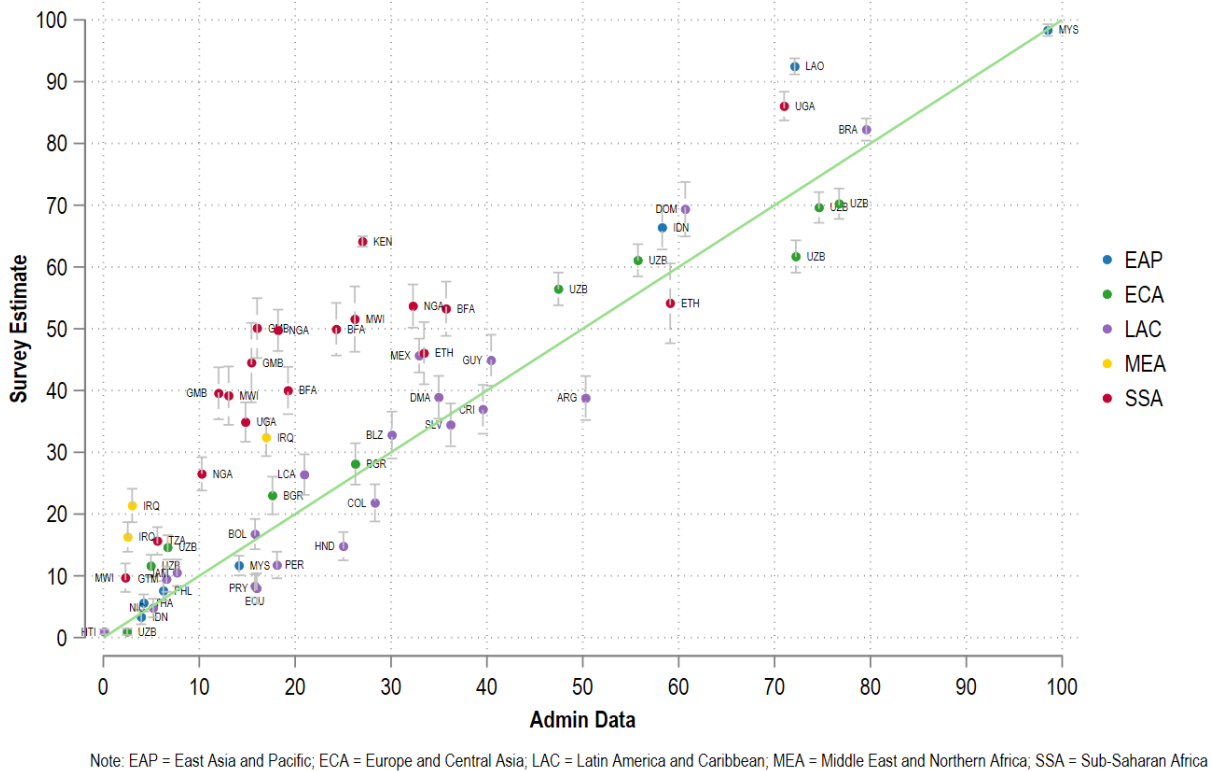
4 Results

4.1 Estimated vaccine coverage in survey and administrative data

We find substantial misalignment between survey-based estimates and administrative records of COVID-19 vaccine coverage. Comparing phone survey estimates to administrative coverage figures from the same time[16] in a sample of 36 LMICs, we find survey estimates that suggest coverage rates that are between 11.5 percentage points lower and 37.1 percentage points higher than those reported in administrative sources (Figure 1). Vaccination rates are statistically different in 28 out of 36 countries (46 out of 57 survey waves). In 21 out of these 28 countries (35 survey waves), survey estimates suggest higher vaccine coverage than is reported in administrative data (in 7 countries survey estimates are lower). On average, survey estimates exceed the administrative data by 47% when excluding extreme positive outliers occurring below 5% reported administrative coverage. However, this pattern displays noticeable regional variation: While estimated survey rates are fairly aligned with administrative figures in Latin America and the Caribbean (LAC), they systematically exceed administrative reports by large margins in Sub-Saharan Africa.

---

[15] Most notably, monotonicity implies that negative reinforcement should not *increase* the propensity of respondents to report being vaccinated and positive reinforcement should not *decrease* their propensity. Furthermore, bounding assumes that the "true" survey-based estimate of the vaccination rate, i.e., the estimate absent any demand/social desirability effects, lies within the bounds provided by the estimated vaccination rate under negative and positive reinforcement. This also has implications for how to choose the framing: Framing should be chosen such that the induced demand effect is *just* stronger (positively as well as negatively) than what is hypothesized to be the case in the original questionnaire design. Inducing a much stronger demand effect leads to demand-robust confidence intervals that are needlessly conservative. In other words: Conditional on inducing a demand effect that is stronger than in the original questionnaire, choosing the question framing that evokes the smallest demand effect is optimal as it leads to the narrowest demand-robust confidence interval within which we expect an experimenter-demand-free survey estimate would lie.

[16] We use the end of the phone survey data collection period as the time of comparison, which allows for some time lag in administrative data reporting.

Figure 1: COVID-19 vaccine coverage estimated from survey data and reported in administrative data

Note: EAP = East Asia and Pacific; ECA = Europe and Central Asia; LAC = Latin America and Caribbean; MEA = Middle East and Northern Africa; SSA = Sub-Saharan Africa

Our subsequent experimental analysis focuses on Sub-Saharan Africa where the observed pattern is most striking. Among the five Sub-Saharan African countries that comprise our experimental sample, phone survey estimates exceed the administrative data by $7 - 32$ percentage points (21% to 320%). The exception is Ethiopia where our in-person survey estimate from May 2022 exceeds the administrative data by 12.6 percentage points (38%) but where a phone survey estimate from January 2023 is statistically indistinguishable from administrative reports (Figure 2).

Figure 2: Estimated COVID-19 vaccine coverage in survey data from five Sub-Saharan African countries vs. administrative records (purposively selected (main) respondents only)
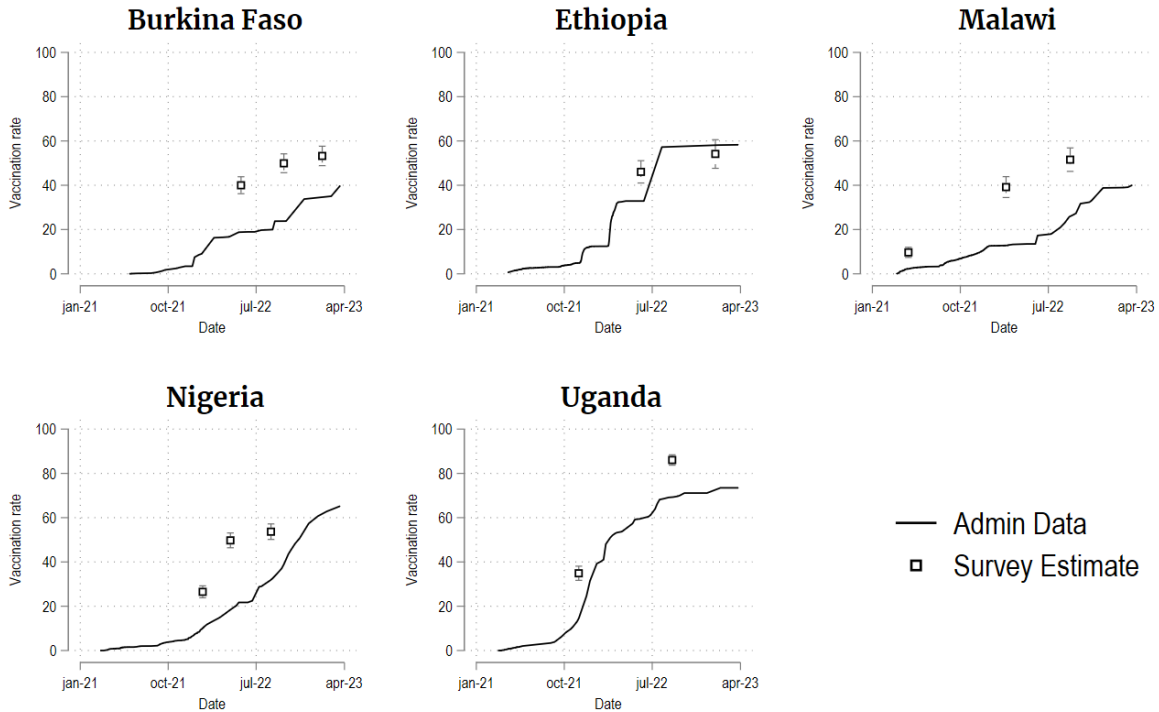
Note: Share of people with at least one COVID-19 vaccine dose assuming no one below 15 was vaccinated. Admin data source: Mathieu et al. (2021)

4.2 Errors of representation

Household sample selection

Our first hypothesis is that the sample of households included in phone surveys is selected and not representative all households in the country. We thus predict that phone survey households would display higher rates of vaccine uptake than households in a nationally representative sample. We test this in data from the Ethiopia Socioeconomic Survey which was collected in-person in May and June of 2022.

Our results do not support this prediction (Table 3). The point estimates of the dummy variable that identifies phone survey households within the full, nationally representative, in-person sample are close to zero and not statistically significant. As a result, there is also little difference between the (weighted) estimate of the national coverage rate using the sample of phone survey households (30.5%) or using the full sample of households included in the Ethiopia Socioeconomic Survey (29.5%). This suggests that selection effects at the household level do not drive the differences in comparison to administrative figures which was reported at 20.1% by the end of in-person data collection.

Table 3: Hypothesis 1 - Household Sample Selection

VARIABLES	(1) Vaccinated
Phone Survey Sample (Dummy)	0.00795
	(0.00523)
Constant	0.272***
	(0.00296)
Observations	33,432
R-squared	0.000
Weighted Vax Rate: ESS Sample	29.5
Weighted Vax Rate: Phone Sample	30.5

Note: Bivariate OLS regression of vaccination status on a dummy for living in a household sampled for the phone survey. The dummy identifies respondents in a general population sample from the Ethiopia Socioeconomic Survey (ESS 5) that live in households that are also included in the phone surveys (vs. the entire general population sample). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Respondent selection

Since the phone surveys' main respondent was purposively selected, we hypothesize that vaccine uptake estimates from the sample of main respondents is biased upwards. We predicted that randomly selecting the household member to be interviewed or collecting (proxy-reported) data on the vaccination status of all household members would lead to lower estimated uptake.

Our results confirm this prediction (Table 4). In all three countries in which we have data on both the main and a randomly selected respondent, estimated uptake is significantly lower among randomly selected respondents compared to purposively selected main respondents. Effect sizes range between five percentage points in Malawi to 13 percentage points in Burkina Faso (Table 4, Panel A). Collecting proxy-reported information for all household members reduces estimated uptake even further, ranging from 13 percentage points in Malawi to 18 percentage points in Uganda (Panel B).

Table 4: Hypothesis 2 - Respondent Selection

	(1)	(2)	(3)
Panel A	**Random vs. Purposive**		
	Burkina Faso	Malawi	Uganda
Random Respondent Selection (Dummy)	-0.131***	-0.0514***	-0.0932***
	(0.0181)	(0.0194)	(0.0126)
Constant	0.469***	0.503***	0.865***
	(0.0118)	(0.0136)	(0.00882)
Observations	2,986	2,655	3,667
R-squared	0.017	0.003	0.015
Panel B	**All adults vs. Purposive**		
All Adults via Proxy Reports (Dummy)	-0.165***	-0.133***	-0.179***
	(0.0128)	(0.0154)	(0.0116)
Constant	0.469***	0.503***	0.865***
	(0.0113)	(0.0133)	(0.0101)
Observations	8,144	5,192	7,770
R-squared	0.020	0.014	0.030
Panel C	**Random vs. All adults**		
Random Respondent Selection (Dummy)	0.0344**	0.0812***	0.0853***
	(0.0142)	(0.0156)	(0.0122)
Constant	0.304***	0.370***	0.687***
	(0.00576)	(0.00785)	(0.00591)
Observations	7,716	5,159	7,693
R-squared	0.001	0.005	0.006
(1) Weighted Vax Rate: Purposively Selected Respondents	49.9	51.5	86.0
(2) Weighted Vax Rate: Randomly Selected Respondents	34.8	45.2	78.0
(3) Weighted Vax Rate: All Household Members 15+ (proxied)	33.3	35.2	68.4
Administrative Data (adult population)	24.3	26.2	71.0
p-value from t-Test of difference, (2) vs. Admin (adult population)	**p < 0.001***	**p < 0.001***	**p < 0.001***
(4) Weighted Vax Rate: All Household Members (proxied)	19.3	20.3	41.0
Administrative Data (general population)	13.6	14.9	38.9
p-value from t-Test of difference, (4) vs. Admin (general population)	**p < 0.001***	**p < 0.001***	**0.010****

Further, there is a statistically significant difference between the estimated vaccination rate based on random respondent selection and based on proxy reports for all household members across all countries (Panel C), which goes against our prediction that both approaches should lead to statistically indistinguishable estimates.

There are several possible explanations for this difference, relating both to errors of representation and errors of measurement. These include errors in proxy reporting (see section 4.2) and enumerator demand (see section 4.3).

After weighting our random respondent estimates using the re-calibrated phone survey weights (see Section 3), the gap between survey data and administrative records remains substantial in all cases (11 percentage points or 46% of administrative coverage in Burkina Faso, 19.5 percentage points or 76% in Malawi, and 8.9 percentage points or 13% in Uganda; Table 4).[17] Estimates are closest to the survey data when using (proxy-reported) data for all household members. However, differences remain statistically significant in all countries and non-negligible in Burkina Faso (5.7 percentage points or 42% of the administrative coverage rate) and Malawi (5.4 percentage points or 36%).

4.3 Measurement errors

Survey mode

We argue that if phone survey mode effects were driving the misalignment between administrative records and survey estimates of vaccine uptake, there should be no misalignment between administrative records and *in-person* survey estimates. We test this by comparing the administrative data to the sample of Ethiopia phone survey respondents who were also interviewed in person in the Ethiopia Socioeconomic Survey (ESS). We find a significant discrepancy of 12.6 percentage points (or 38%) between administrative records (33.4% vaccine coverage)[18] and in-person survey estimates (46.0% coverage) in this sample (Table 5). This difference remains substantial at 10.4 percentage points (or 52%) and statistically significant when estimating vaccine coverage based on the full ESS sample, which is representative of the general population. Survey

---

[17] At the time of data collection, Malawi and Uganda had already been vaccinating children under the age of 16 for over 8 and over 13 months, respectively. This means that our assumption of zero vaccinated children under the age of 15 that we have to make for comparability between survey data estimates for the adult population (15+) and administrative data is likely too conservative and understates the true gap in this case.

[18] Conservatively assuming that at the time of reporting no vaccinated children were included in the administrative data.

mode effects thus do not appear to be driving the observed difference between survey and administrative data.

Table 5: Hypothesis 3 - Survey Mode

Panel A: Interviewing phone survey main respondents in-person

Phone Survey Main Respondents, Apr - Jun 2022	46.0
	(41.0 to 51.1)
Admin data (adults 15+), June 5	33.4
N (Phone Survey Main Respondents)	2,403
t-Test, Main respondents (survey) vs. Admin (adults 15+)	p < 0.0001 ***

Panel B: Collecting vaccination status of all phone survey household members in-person

All Phone Survey Household Members, Apr - Jun 2022	30.5
	(27.2 to 33.7)
Admin data (general population), June 5	20.1
N (All Phone Survey Household Members)	10,786
t-Test, All HH members (survey) vs. Admin (general population)	0.0001 ***

Note: Weighted estimates of vaccine uptake in face-to-face data from the Ethiopia Socioeconomic Survey, Wave 5 (ESS 5) (Apr-June 2022). Administrative vaccine coverage rate for adults conservatively assumes that no children below the age of 15 were vaccinated at the time of reporting. All values in percent. 95%-Confidence Intervals in parentheses.

Panel conditioning

Respondents may change their vaccination behavior in response to being repeatedly surveyed on the topic of COVID-19 vaccination. We therefore predicted that respondents previously interviewed or those in households that were part of the phone survey panel would report higher rates of vaccination than a sample interviewed for the first time.

Table 6: Hypothesis 4 - Panel Conditioning

VARIABLES	(1) Panel Household	(2) Panel Individual
Household Previously Interviewed	-0.0177	
	(0.0172)	
Respondent Previously Interviewed		0.00847
		(0.0172)
Constant	0.324***	0.310***
	(0.0129)	(0.0117)
Observations	2,942	2,942
R-squared	0.000	0.000

Note: Bivariate OLS regression of vaccination status on a dummy for respondent living in a household interviewed before (Column 1) or having been interviewed before themselves (Column 2). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We do not find support for any effects of previous survey participation on reported vaccine uptake (Table 6). Point estimates are close to zero and not statistically significant.

Proxy reporting

Earlier, we found that collecting the vaccination status of all household members via proxy reporting led to estimates of vaccine coverage that were closest to the coverage suggested in administrative data. We suspected that part of the remaining difference may be explained by the accuracy of proxy reports, that is, cases in which the reporting main respondent overstated the vaccination status of a household member.

Table 7: Hypothesis 5 - Proxy Reporting

VARIABLES	(1) Burkina Faso	(2) Malawi	(3) Uganda
Self reporting (Dummy)	0.0396*	0.110***	0.0267
	(0.0219)	(0.0256)	(0.0211)
Constant	0.244***	0.309***	0.668***
	(0.0155)	(0.0182)	(0.0149)
Observations	1,616	1,394	1,950
R-squared	0.002	0.013	0.001

Note: Bivariate OLS regression of vaccination status on a dummy for self reporting (as opposed to proxy reporting). Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

When comparing self- and proxy-reported vaccination status for the same individuals, we find the opposite to be the case (Table 7). Proxy reports prove fairly aligned with self-reports, coinciding in 94% of cases in Burkina Faso, 81% of cases in Malawi, and 93% of cases in Uganda (Table A2). However, proxy reports tend to *under*state vaccine uptake compared to self-reports of the same individuals, statistically significantly so in Burkina Faso and Malawi. Arguably, it is plausible that main respondents fail to observe or remember the vaccination of members of their households. This implies that when using self-reports wherever available and proxy reports otherwise, the gap between survey and administrative coverage rates would increase slightly (Figure 3).

We also collect proxy reports for the number of doses received and, for those reported to not be vaccinated yet, willingness to get vaccinated (Table A2). We find that proxy reporting still provides information that is consistent with self-reports in the case of the number of doses received (96% in Burkina Faso, 70% in Malawi, 88% in Uganda) but is essentially as good as a random guess when asking for vaccine acceptance (53% agreement in Burkina Faso, 56% in Malawi, 54% in Uganda).

Experimenter demand

Our analysis so far has treated self-reported information as accurate. However, one's own COVID-19 vaccination status may be a sensitive topic, inducing respondents to exaggerate vaccine uptake in response to "experimenter demand" effects.[19]

We find no support for such behavior (Table 8). When making experimenter demand effects explicit by framing the vaccination uptake question with a positive or negative expectation from the enumerator, we obtain near identical and statistically indistinguishable estimates of vaccine uptake. Furthermore, negative and positive reinforcement of experimenter demand produces estimates that are statistically indistinguishable from the standard question framing. We take the fact that the estimates are remarkably close to those under our standard question framing even when explicitly introducing experimenter demand to indicate the robustness of self-reported vaccine uptake to such effects.

Table 8: Hypothesis 6 - Experimenter Demand

VARIABLES	(1) Burkina Faso	(2) Ethiopia
Treatment Arm = 1, Positive demand	0.0137	0.0146
	(0.0300)	(0.0243)
Treatment Arm = 2, Negative demand	0.0155	0.0284
	(0.0301)	(0.0245)
Constant	0.487***	0.475***
	(0.0214)	(0.0171)
Observations	1,668	2,509
R-squared	0.000	0.001

Note: OLS regression of vaccination status on dummies for treatment group 1 (positive demand) and treatment group 2 (negative demand). Base category is standard phrasing. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

4.4 Misreporting in administrative data

So far, our analysis has found that errors of representation in the survey data explain part but not all of the gap between survey estimates and administrative figures. We next turn to possible sources of error in the administrative data. While we cannot experimentally probe the administrative data in the same way as the survey data, we present indicative evidence pointing to some weaknesses and inaccuracies in administrative data.

---

[19] This pattern would also be consistent with self-reported information providing higher estimates of vaccine uptake than proxy-reported information.

Weak administrative data systems

Previous literature has indicated that weak data systems may give rise to inaccurate administrative statistics in the case of routine immunization data in LMICs (Cutts et al. 2016; Sandefur and Glassman 2015). This may also be the case for reported COVID-19 vaccination rates. We investigate whether there is a correlation between the size of the gap between survey and administrative data (measured as their absolute percent difference) and countries' statistical capacity (measured by the World Bank's Statistical Performance Indicators (SPI)).

Table 9: Statistical performance and survey-admin gap in vaccination rates

VARIABLES	(1) Gap Survey vs. Admin	(2) Gap Survey vs. Admin	(3) Gap Survey vs. Admin
SPI: Total score	-1.608**	-1.187	-0.990
	(0.743)	(0.770)	(0.776)
SPI: Dimension 4.2 score (Admin data)	-55.37***	-46.28***	1.160
	(15.36)	(15.75)	(9.335)
GDP per capita (PPP 100K)		-0.180	0.0321
		(0.108)	(0.0955)
Region Code = 2, ECA			4.716
			(13.76)
Region Code = 3, LCN			4.613
			(7.408)
Region Code = 4, MEA			41.25
			(28.46)
Region Code = 7, SSA			103.1***
			(18.60)
Constant	182.0***	167.0***	78.99*
	(47.69)	(46.97)	(44.66)
Observations	48	48	48
R-squared	0.345	0.364	0.603

Note: Results from OLS regressions of the absolute gap between survey estimates and administrative figures of vaccination coverage on indicators of statistical performance. To avoid extreme leverage at low values of coverage, the sample is constrained to those survey waves where administrative data coverage was at least 5% of the adult population. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We find lower statistical performance scores to be significantly associated with higher gaps between the survey and administrative data in our sample of LMICs (Table 9). In a bivariate regression, the coefficient estimate suggests that a 1 percentage point increase in the total SPI score is associated with a 1.6 percentage point reduction in the percent gap between both data sources. Further, the availability of Civil Registration and Vital Statistics (CRVS), the SPI's metric for administrative data quality specifically, is associated with a 55.4 percentage point smaller gap. When additionally controlling for GDP per capita, we find that only the SPI indicator of administrative data quality is significantly and negatively associated with the percent gap between

survey and administrative vaccination figures. When additionally controlling for region fixed effects, the correlation between statistical performance and discrepancies in vaccination rates is no longer statistically significant with the effect absorbed by the Sub-Saharan Africa regional dummy.

We take this as evidence that discrepancies between administrative records and survey data are more likely and larger where administrative data systems are weaker and suffer from lower capacity.

Numerator issues (number of people currently vaccinated)

Next, we turn our attention to the possible sources of bias in the administrative data (Table 10). We distinguish between numerator issues (misreporting of the number of people currently vaccinated) and denominator issues (under- or overestimating target population size).

On the side of potential numerator issues, we find large time gaps in reporting to be common. While the average frequency of reporting is fairly high (every 2.4 days) across our sample of 36 LMICs, gaps can get as long as close to half a year without a vaccination rate report (175 days between July 2022 and January 2023 in Ethiopia). The longest time gaps between two data points are almost two months (54.9 days) when averaging across our sample of countries, with 26 out of 36 countries (72%) having gaps of 30 days or longer. Fourteen countries (42%) have gaps longer than 2 months (60 days). Gaps are even greater in our sample of five Sub-Saharan African countries that are the subject of our survey experiments. Here, the average longest gap amounts to 3 months (90 days) with all countries displaying gaps between reports of 42 days or more.

We also find large jumps in reported coverage between data points in the administrative data. Across our sample of LMICs, as well as in our smaller sample of five Sub-Saharan African countries, we find the average increase in reported COVID-19 vaccine coverage to amount to 0.2 percentage points in between two reports (0.11pp and 0.04pp on a per-day basis, respectively). However, these small increases on average hide some larger jumps. In 16 out of 36 countries (44%), jumps of 5 percentage points or more occur in between reports and in seven countries, jumps exceed 10 percentage points. There are extreme cases, such as in Nicaragua where reported coverage jumped by 37 percentage points within a two-week window in November 2021 or Guyana that reported vaccinating 4% of its population on a single day in November 2022. Owing to the slow progress of vaccination campaigns in the region, these jumps are somewhat smaller but still substantial in our sample of Sub-Saharan African countries.

Given the at times infrequent reporting of vaccination rates and large jumps in coverage, one possibility would be that cases where survey estimates exceed administrative figures reflect lags in the administrative data. We bound these potential lags by reporting the mean number of days it takes for administrative coverage to catch up with coverage estimated from surveys wherever survey estimates exceeded administrative reports to start with among the 36 LMICs we study.[20] We find that these lags would have to be very substantial in order for them to explain the

---

[20] As such, the analysis sample for this exercise excludes cases where survey estimates are below administrative reports and where the administrative data had not caught up to survey estimates by March 31, 2023.

discrepancy with higher survey estimates. On average, over three months (97.9 days) pass until administrative reports reach the coverage estimated from our baseline survey data estimates. These lags would have to exceed six months (182 days) on average in our sample of five Sub-Saharan African countries. Of note, they would still have to be substantial even when we account for respondent selection effects in the survey data (38 days in Uganda, 63 days in Burkina Faso, 102 days in Malawi).

Denominator issues (under- or overestimating target population size)

We also find evidence for denominator issues in the administrative data (Table 10). To analyze this, we compare administrative coverage figures, expressed in percent of the total population, between the Our World in Data (OWID) COVID-19 Vaccination data set (the administrative data source for our study, Mathieu et al. 2021) and the WHO's COVID-19 vaccination dashboard (WHO 2020b). In most cases, the total number of people vaccinated at a given date coincides exactly between both data source. This allows us to compare reported coverage as a total population share for the same date and for the same figure of total number of people vaccinated in the OWID and WHO data. Any discrepancies we detect between vaccination rates (expressed in percent of the population) thus reflect differences in the denominator, that is, the size of the population. We can make this comparison for 19 LMICs.

Table 10: Potential Numerator and Denominator Issues in Administrative Data

Issue	Metric	Countries	Mean	Max	Min
Numerator Issues (misreporting of the number of people currently vaccinated)	Average frequency of reporting	36 LMICs	2.4	175	1
		5 LMICs in SSA	4.5	175	1
	Average longest time gap between two data points	36 LMICs	54.9	175	1
		5 LMICs in SSA	90.4	175	42
	Largest jump in administrative data (percentage points)	36 LMICs	0.20	37.3	0
		5 LMICs in SSA	0.20	11.10	0
	Largest jump in administrative data (percentage points per day since last report)	36 LMICs	0.11	4.19	0
		5 LMICs in SSA	0.04	1.33	0
	Average time lag for admin data to catch up with survey estimate	22 LMICs	97.9	442	1
		4 LMICs in SSA	182.6	308	35
Denominator Issues (under- or overestimation of target population size)	Average difference in reported vaccination rate for identical number of people covered: WHO vs. Our World in Data	19 LMICs with exact match in date and number of people vaccinated	1.6	-2.9	5.1
		5 LMICs in SSA	1.9	1.4	2.6

We find that the discrepancies introduced by denominator issues are non-negligible and amount to a difference of up to 5 percentage points between coverage rates reported in the WHO dashboard and the OWID dataset. On average, reported coverage in the WHO dashboard is 1.6 percentage points higher than that reported in the OWID data. Similarly, coverage reported in the WHO dashboard is on average 1.9 percentage points higher in our sample of five Sub-Saharan African countries. This implies that, on average, the coverage rates reported in the WHO data assume a smaller population size. However, cases also exist in which assumed population size in the WHO data is higher, leading to up to 2.9 percentage points lower vaccination rates.

These discrepancies seem to arise because of differences in baseline years for which population figures are taken. The OWID dataset uses the latest, 2022 UN population projections (Mathieu et al. 2021), whereas population figures implied by the WHO's reported coverage rates typically correspond to UN World Population Prospects for 2019 or 2020.

5 Discussion

Phone surveys have become commonplace during the COVID-19 pandemic and have informed a large body of (health) policy research. Our results from 57 survey waves on COVID-19 vaccination across 36 LMICs indicate that phone surveys may produce estimates of COVID-19 vaccine coverage that differ from administrative figures. In some cases, survey estimates exceed administrative figures by a margin that would suggest vastly different policy conclusions. Such misalignment is particularly concentrated in Sub-Saharan Africa while coverage rates estimated in other regions on average track administrative records more closely.

Upon investigating potential sources of bias, our findings largely maintain the reliability of the phone survey data. We find phone survey estimates from five Sub-Saharan African countries to be robust to a number of commonly feared representation and measurement errors, but they are affected by how respondents are selected (Figure 3 and Appendix Table A3).

We find little evidence that errors of measurement bias the phone survey estimates. Interviewing respondents in person rather than on the phone does not lead to survey estimates that are aligned with administrative records. Similarly, reported vaccine uptake does not differ significantly between first-time and previous respondents of the phone surveys suggesting that panel conditioning effects do not bias estimates. Contrary to Wolter et al. (2022) who use a list experiment in an online sample in Germany to find experimenter demand biasing reported vaccine uptake, we do not find evidence of such effects in our sample. While we find proxy reports to have generally good accuracy, we do find that they miss out on some vaccinations obtained by other household members.

As for sample selection, we find no statistically significant difference in estimated vaccination rates between the phone survey sample of households and the full nationally representative sample of households. However, respondent selection matters. In line with the results of Bradley et al. (2021) from online surveys in the United States, estimated vaccine uptake is significantly lower when taking into account respondent selection effects at the individual level (Figure 3). Moving

from interviewing a purposively selected household member that is knowledgeable about the affairs of the household, a common approach in multi-topic phone surveys (Gourlay et al. 2021), to randomly selecting the respondent reduces estimated vaccine uptake by 10 percentage points on average. This is equivalent to a reduction of the gap between survey estimates and administrative figures by 37% on average. We find that collecting proxy-reported information on all household members has a similar, or even stronger, effect and reasonable accuracy even though proxy reports seem to miss out on some vaccinations.

There is thus a trade-off between selecting a respondent that is knowledgeable across a broad domain of topics and obtaining representative results on a single, individual-level metric such as vaccine uptake, attitudes, or personal sentiments. If the latter are the topic of interest, our results show that random respondent selection, wherever possible from an up-to-date roster of household members, can successfully mitigate sample selection issues. For multi-topic surveys and where the information of interest is sufficiently salient to other household members, collecting proxy-reported information through a knowledgeable, purposively selected household member can constitute an alternative that is reasonably accurate and less challenging to implement from a survey design perspective.

Figure 3: Estimated COVID-19 vaccine coverage under varying survey design choices
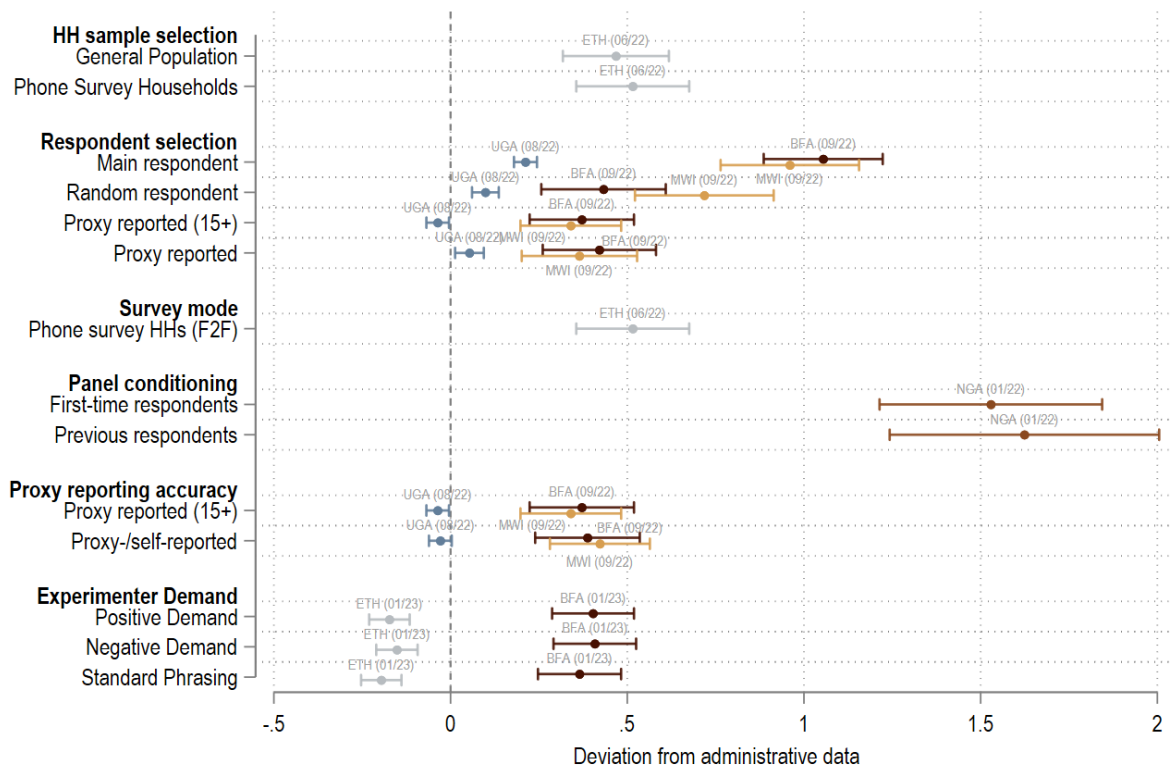
Finally, discrepancies between survey estimates of COVID-19 vaccine coverage and administrative data may also be related to flaws and weaknesses in administrative data systems (Shapira et al. 2021; Rosenbaum and Waugaman 2022). Gaps in the administrative data can, for

example, arise when handwritten records are not fully digitized, reported with error, or records are not passed on by all health care providers (Shapira et al. 2021). In the case of COVID-19, evidence from a survey of 42 USAID country offices has documented that such issues have been commonplace (Rosenbaum and Waugaman 2022).

We provide indicative evidence of such flaws and weaknesses, which had previously also been documented in the case of routine immunizations (Lim et al. 2008; Sandefur and Glassman 2015; Cutts et al. 2016; Galles et al. 2021) and quantify their potential effects on reported COVID-19 vaccination rates. Survey data can thus serve as a complementary source to countercheck the official administrative figures and inform policy.

To this end, phone surveys can provide flexible, cost-effective, and rapidly deployable tools for high-frequency monitoring. Phone surveys also facilitate iterative experimentation with design choices to ascertain the robustness of the evidence compiled. This would either be prohibitively expensive or outright impossible at similar scale in the case of in-person surveys or administrative data.

Both monitoring and experimentation was possible because of the existing survey infrastructure building on longitudinal in-person households, which allowed for broad population coverage and comparatively low non-response rates (Gourlay et al. 2021) as well as providing a nationally representative benchmark for the phone surveys.

6 Conclusion

Studies from before the COVID-19 pandemic have found substantial discrepancies in LMICs between routine childhood immunization rates reported in survey and administrative data, hence creating a dilemma for research and policy (Galles et al. 2021; Danovaro-Holliday et al. 2021; Cutts et al. 2016; Burton et al. 2009; Lim et al. 2008). In this study, we show that such misalignment is also widespread in the context of COVID-19 vaccinations. In a sample of 36 LMICs, the survey data statistically significantly exceeds figures reported in administrative sources in 35 out of 57 survey rounds and by 47% on average. This pattern is particularly striking and consistent in Sub-Saharan Africa.

We investigate a number of potential explanations for this discrepancy, focusing on possible errors of representation and errors of measurement in phone survey data from five Sub-Saharan African countries. The gap between both data sources shrinks when accounting for selection bias at the respondent level but remains substantial in most cases. While we cannot experimentally probe the administrative data in the same way as the survey data, we present indicative evidence pointing to some weaknesses and inaccuracies in administrative data.

Our findings make several substantial contributions.

First, we show that substantial misalignment between survey-based and administrative vaccine coverage rates plagues COVID-19 data, the largest vaccination effort in history. Our study is the first to document these discrepancies in a cross-country sample and a low- and lower-middle

income context. We show that the direction of misalignment generally runs counter to what was previously observed in the case of routine immunization rates before the pandemic: Survey estimates systematically exceed vaccine coverage reported in administrative sources across most countries we study. This evidence suggests that the ongoing effort to reach widespread COVID-19 immunization in LMICs relies on data that would imply (sometimes vastly) different policy conclusions depending on the data source consulted.

Second, our results suggest that phone surveys can be a suitable and reliable tool for COVID-19 vaccination research and possibly other health policy contexts, supporting the robustness of findings from a large body of research during COVID-19 that relied on these data (Kanyanda et al. 2021; Solís Arce et al. 2021; A. Reza et al. 2022; Wollburg, Markhof, et al. 2022; Wollburg et al. 2023). While survey estimates are somewhat sensitive to design choices, especially the selection of respondents to be interviewed, commonly feared and hard to ascertain non-sampling errors in phone surveys, such as mode effects, panel conditioning, and experimenter demand, appear not to affect estimates of vaccine uptake meaningfully. Careful survey design can limit the effects of sampling errors.

Third, our experimental results inform best practices for future phone survey-based research and policy beyond the COVID-19 context. Our results suggest that random respondent selection, wherever possible from an up-to-date roster of household members, can successfully mitigate errors of representation and should be preferred for collecting detailed, individual-level information on vaccine and health-related issues. This particularly applies where the information collected would be less salient to other household members such as (vaccine) attitudes and personal sentiments. Collecting proxy-reported information through a knowledgeable, purposively selected household member can constitute an alternative that is reasonably accurate for salient information such as vaccine uptake, is less challenging to implement from a survey design perspective, and is applicable to multi-topic surveys that require the selection of a broadly knowledgeable respondent.

In this sense, our findings also underscore the complementarity between phone and face-to-face surveys. The latter remain indispensable to provide a recent sampling frame for households and individuals which our results suggest are some of the mainstays of reliable estimates. At the same time, phone survey data can complement face-to-face data collection. For example, the usefulness of survey data for policy has been limited by its low frequency (every 2-3+ years) and the difficulty of conducting surveys in hard-to-access or conflict-affected areas (Cutts et al. 2016, Danovaro-Holiday et al. 2021, Burton et al. 2009). Phone surveys offer the opportunity to improve on these limitations.

As phone surveys have been proposed as flexible vehicles for (health) data collection (Gourlay et al. 2021, Zezza et al. 2022, Glazerman et al. 2023) our findings may find direct application in future (vaccine) data collection efforts. Such efforts may include campaigns to catch up on (foregone) child immunization which saw the largest sustained decline in 30 years during the COVID-19 pandemic (WHO 2022, The Lancet Child & Adolescent Health 2021, Causey et al. 2021) or a large-scale rollout of the first Malaria vaccine (Wilyard 2022).

Our study faces several limitations. Most notably, we cannot assess the reliability of administrative data directly. Our focus on survey-based sources of error is justified due to the importance of survey data for health research and policy but leaves the possibility of (non-negligible) error in the administrative data (see Sandefur and Glassman 2015, Lim et al. 2008). In this regard, the evidence we present on potential issues in administrative data sources should only be considered indicative. Having access to disaggregated administrative data, for example by sex, age, and lower administrative levels, or gaining systematic insights into the data pipeline would allow for a more complete assessment of the matter. Similarly, we cannot benchmark our estimates against an objective truth that would allow us to exactly quantify the amount of bias in either data source or collect self-reported information from each household member. Our study is limited to assessing the relative bias of different design choices. We further cannot rule out that there are interaction effects between different sources of potential bias, such as between survey mode and other sources of error. While most of our experiments were conducted across multiple countries, we cannot ascertain whether they would in all cases provide the same results when repeated for a different set of countries or at a different time.

Returning to this paper's title, our study suggests that for vaccination campaigns in LMICs to be accurately informed, researchers and policy makers should pay close attention to discrepancies between different sources of data and the possibly diverging policy conclusions they purport. This study advances our understanding of the sources of such discrepancies in (phone) surveys and concludes that when adequately designed, they can become an important asset in the health data collection toolkit of researchers and policy makers alike.

Bibliography

Adhvaryu, Achyuta, James Fenske, and Anant Nyshadham. 2019. "Early Life Circumstance and Adult Mental Health." *Journal of Political Economy* 127 (4): 1516–49. https://doi.org/10.1086/701606.

Aggarwal, Shilpa. 2021. "The Long Road to Health: Healthcare Utilization Impacts of a Road Pavement Policy in Rural India." *Journal of Development Economics* 151 (June): 102667. https://doi.org/10.1016/j.jdeveco.2021.102667.

Althubaiti, Alaa. 2016. "Information Bias in Health Research: Definition, Pitfalls, and Adjustment Methods." *Journal of Multidisciplinary Healthcare*, May, 211. https://doi.org/10.2147/JMDH.S104807.

Amaya, Ashley, Paul P Biemer, and David Kinyon. 2020. "Total Error in a Big Data World: Adapting the TSE Framework to Big Data." *Journal of Survey Statistics and Methodology* 8 (1): 89–119. https://doi.org/10.1093/jssam/smz056.

Ambel, Alemayehu, Kevin McGee, and Asmelash Tsegay. 2021. "Reducing Bias in Phone Survey Samples. Effectiveness of Reweighting Techniques Using Face-to-Face Surveys as Frames in Four African Countries." 9676. Policy Research Working Paper. Washington D.C.: World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/859261622035611710/reducing-bias-in-phone-survey-samples-effectiveness-of-reweighting-techniques-using-face-to-face-surveys-as-frames-in-four-african-countries.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2019. "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials." *The Review of Economic Studies* 86 (6): 2453–90. https://doi.org/10.1093/restud/rdz008.

Banerjee, Abhijit, E. Duflo, R. Glennerster, and D. Kothari. 2010. "Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives." *BMJ* 340 (may17 1): c2220–c2220. https://doi.org/10.1136/bmj.c2220.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *The World Bank Economic Review* 25 (3): 418–47. https://doi.org/10.1093/wber/lhr022.

Barham, Tania, and John A. Maluccio. 2009. "Eradicating Diseases: The Effect of Conditional Cash Transfers on Vaccination Coverage in Rural Nicaragua." *Journal of Health Economics* 28 (3): 611–21. https://doi.org/10.1016/j.jhealeco.2008.12.010.

Beaman, Lori, and Andrew Dillon. 2012. "Do Household Definitions Matter in Survey Design? Results from a Randomized Survey Experiment in Mali." *Journal of Development Economics* 98 (1): 124–35. https://doi.org/10.1016/j.jdeveco.2011.06.005.

Benedetti, Fiorella, Pablo Ibarrarán, and Patrick J. McEwan. 2016. "Do Education and Health Conditions Matter in a Large Cash Transfer? Evidence from a Honduran Experiment." *Economic Development and Cultural Change* 64 (4): 759–93. https://doi.org/10.1086/686583.

Biemer, P. P. 2010. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74 (5): 817–48. https://doi.org/10.1093/poq/nfq058.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124 (2): 735–69. https://doi.org/10.1162/qjec.2009.124.2.735.

Blimpo, Moussa P., Pedro Carneiro, Pamela Jervis, and Todd Pugatch. 2022. "Improving Access and Quality in Early Childhood Development Programs: Experimental Evidence from the Gambia."

Economic Development and Cultural Change 70 (4): 1479–1529. https://doi.org/10.1086/714013.

Bradley, Valerie C., Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng, and Seth Flaxman. 2021. "Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake." *Nature* 600 (7890): 695–700. https://doi.org/10.1038/s41586-021-04198-4.

Brown, David W., Anthony Burton, Marta Gacic-Dobo, and Rouslan Karimov. 2013. "An Introduction to the Grade of Confidence Used to Characterize Uncertainty Around the WHO and UNICEF Estimates of National Immunization Coverage." *The Open Public Health Journal* 6 (1): 73–76. https://doi.org/10.2174/1874944501306010073.

Brubaker, Joshua, Talip Kilic, and Philip Wollburg. 2021. "Representativeness of Individual-Level Data in COVID-19 Phone Surveys: Findings from Sub-Saharan Africa." Edited by Bjorn Van Campenhout. *PLOS ONE* 16 (11): e0258877. https://doi.org/10.1371/journal.pone.0258877.

Burton, Anthony, Robert Kowalski, Marta Gacic-Dobo, Rouslan Karimov, and David Brown. 2012. "A Formal Representation of the WHO and UNICEF Estimates of National Immunization Coverage: A Computational Logic Approach." Edited by Thomas Eisele. *PLoS ONE* 7 (10): e47806. https://doi.org/10.1371/journal.pone.0047806.

Burton, Anthony, Roeland Monasch, Barbara Lautenbach, Marta Gacic-Dobo, Maryanne Neill, Rouslan Karimov, Lara Wolfson, Gareth Jones, and Maureen Birmingham. 2009. "WHO and UNICEF Estimates of National Infant Immunization Coverage: Methods and Processes." *Bulletin of the World Health Organization* 87 (7): 535–41. https://doi.org/10.2471/BLT.08.053819.

Celhay, Pablo A., Julia Johannsen, Sebastian Martinez, and Cecilia Vidal. 2021. "Can Small Incentives Have Large Payoffs? Health Impacts of a Cash Transfer Program in Bolivia." *Economic Development and Cultural Change* 69 (2): 591–621. https://doi.org/10.1086/703085.

Chandir, Subhash, Danya Arif Siddiqi, Sara Abdullah, Esther Duflo, Aamir Javed Khan, and Rachel Glennerster. 2022. "Small Mobile Conditional Cash Transfers (MCCTs) of Different Amounts, Schedules and Design to Improve Routine Childhood Immunization Coverage and Timeliness of Children Aged 0-23 Months in Pakistan: An Open Label Multi-Arm Randomized Controlled Trial." *EClinicalMedicine* 50 (August): 101500. https://doi.org/10.1016/j.eclinm.2022.101500.

Christensen, Darin, Oeindrila Dube, Johannes Haushofer, Bilal Siddiqi, and Maarten Voors. 2021. "Building Resilient Health Systems: Experimental Evidence from Sierra Leone and The 2014 Ebola Outbreak." *The Quarterly Journal of Economics* 136 (2): 1145–98. https://doi.org/10.1093/qje/qjaa039.

Cockx, Lara. 2022. "Moving toward a Better Future? Migration and Children's Health and Education." *Economic Development and Cultural Change* 70 (3): 1229–93. https://doi.org/10.1086/713931.

Cutts, Felicity T., Pierre Claquin, M. Carolina Danovaro-Holliday, and Dale A. Rhoda. 2016. "Monitoring Vaccination Coverage: Defining the Role of Surveys." *Vaccine* 34 (35): 4103–9. https://doi.org/10.1016/j.vaccine.2016.06.053.

Dang, Hai-Anh H., John Pullinger, Umar Serajuddin, and Brian Stacy. 2023. "Statistical Performance Indicators and Index—a New Tool to Measure Country Statistical Capacity." *Scientific Data* 10 (1): 146.

Danovaro-Holliday, M Carolina, Katrina Kretsinger, and Marta Gacic-Dobo. 2021. "Measuring and Ensuring Routine Childhood Vaccination Coverage." *The Lancet* 398 (10299): 468–69. https://doi.org/10.1016/S0140-6736(21)01228-9.

Dansereau, Emily, David Brown, Lena Stashko, and M. Carolina Danovaro-Holliday. 2020. "A Systematic Review of the Agreement of Recall, Home-Based Records, Facility Records, BCG Scar, and Serology for Ascertaining Vaccination Status in Low and Middle-Income Countries." *Gates Open Research* 3 (February): 923. https://doi.org/10.12688/gatesopenres.12916.2.

Das, Jishnu, Jeffrey Hammer, and Carolina Sánchez-Paramo. 2012. "The Impact of Recall Periods on Reported Morbidity and Health Seeking Behavior." *Journal of Development Economics* 98 (1): 76–88. https://doi.org/10.1016/j.jdeveco.2011.07.001.

Davin, Bérengère, Xavier Joutard, and Alain Paraponaris. 2019. ""If You Were Me"': Proxy Respondents' Biases in Population Health Surveys." WP 2019-Nr 05. Aix-Marseille School of Economics.

Dayton, Julia M., Ifeanyi Nzegwu Edochie, David Locke Newhouse, Alexandru Cojocaru, Gildas Bopahbe Deudibe, Jakub Jan Kakietek, Yeon Soo Kim, and Jose Montes. 2022. "COVID-19 Vaccine Hesitancy in 53 Developing Countries : Levels, Trends, and Reasons for Hesitancy." *Policy Research Working Paper Series*, Policy Research Working Paper Series, , September. https://ideas.repec.org/p/wbk/wbrwps/10191.html.

De, Prabal K., and Laxman Timilsina. 2020. "Cash-based Maternal Health Interventions Can Improve Childhood Vaccination—Evidence from India." *Health Economics* 29 (10): 1202–19. https://doi.org/10.1002/hec.4129.

De Weerdt, Joachim, John Gibson, and Kathleen Beegle. 2020a. "What Can We Learn from Experimenting with Survey Methods?" *Annual Review of Resource Economics* 12 (1): 431–47. https://doi.org/10.1146/annurev-resource-103019-105958.

———. 2020b. "What Can We Learn from Experimenting with Survey Methods?" *Annual Review of Resource Economics* 12 (1): 431–47. https://doi.org/10.1146/annurev-resource-103019-105958.

Debnath, Sisir. 2021. "Improving Maternal Health Using Incentives for Mothers and Health Care Workers: Evidence from India." *Economic Development and Cultural Change* 69 (2): 685–725. https://doi.org/10.1086/703083.

Dillon, Andrew, Steven Glazerman, and Mike Rosenbaum. 2021. "Understanding Response Rates in Random Digit Dial Surveys." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3836024.

Dillon, Andrew, Dean Karlan, Christopher Udry, and Jonathan Zinman. 2020. "Good Identification, Meet Good Data." *World Development* 127 (March): 104796. https://doi.org/10.1016/j.worlddev.2019.104796.

Dykstra, Sarah, Amanda Glassman, Charles Kenny, and Justin Sandefur. 2019. "Regression Discontinuity Analysis of Gavi's Impact on Vaccination Rates." *Journal of Development Economics* 140 (September): 12–25. https://doi.org/10.1016/j.jdeveco.2019.04.005.

Galles, Natalie C, Patrick Y Liu, Rachel L Updike, Nancy Fullman, Jason Nguyen, Sam Rolfe, Alyssa N Sbarra, et al. 2021. "Measuring Routine Childhood Vaccination Coverage in 204 Countries and Territories, 1980–2019: A Systematic Analysis for the Global Burden of Disease Study 2020, Release 1." *The Lancet* 398 (10299): 503–21. https://doi.org/10.1016/S0140-6736(21)00984-3.

Gavi, The Vaccine Alliance. 2021. "Zero-Dose Children and Missed Communities." Gavi, the Vaccine Alliance. November 4, 2021. https://www.gavi.org/our-alliance/strategy/phase-5-2021-2025/equity-goal/zero-dose-children-missed-communities.

———. 2022a. "Gavi Vaccine Funding Guidelines." Gavi, The Vaccine Alliance. https://www.gavi.org/sites/default/files/document/2022/Vaccine_FundingGuidelines_0.pdf.

———. 2022b. "Gavi Application Process Guidelines." Gavi, The Vaccine Alliance. https://www.gavi.org/sites/default/files/support/ApplicationProcess_Guidelines.pdf.

———. 2022c. "Data." Gavi, the Vaccine Alliance. November 29, 2022. https://www.gavi.org/types-support/health-system-and-immunisation-strengthening/data.

Glazerman, Steven, Karen A. Grépin, Valerie Mueller, Michael Rosenbaum, and Nicole Wu. 2023. "Do Referrals Improve the Representation of Women in Mobile Phone Surveys?" *Journal of Development Economics* 162 (May): 103077. https://doi.org/10.1016/j.jdeveco.2023.103077.

Glazerman, Steven, Michael Rosenbaum, Rosemarie Sandino, and Lindsey Shaughnessy. 2020. "Remote Surveying in a Pandemic: Handbook." Innovation for Poverty Action. https://www.poverty-action.org/sites/default/files/publications/IPA-Phone-Surveying-in-a-Pandemic-Handbook.pdf.

Gourlay, Sydney, Talip Kilic, Antonio Martuscelli, Philip Wollburg, and Alberto Zezza. 2021. "Viewpoint: High-Frequency Phone Surveys on COVID-19: Good Practices, Open Questions." *Food Policy* 105 (December): 102153. https://doi.org/10.1016/j.foodpol.2021.102153.

Groves, Robert M. 1989. *Survey Errors and Survey Costs: Groves/Survey Errors*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471725277.

Groves, Robert M., and L. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74 (5): 849–79. https://doi.org/10.1093/poq/nfq065.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: ExperimentalEvidence from Kenya." *The Quarterly Journal of Economics* 131 (4): 1973–2042. https://doi.org/10.1093/qje/qjw025.

Henderson, Savanna, and Michael Rosenbaum. 2020. "Remote Surveying in a Pandemic: Research Synthesis." Innovation for Poverty Action. http://www.poverty-action.org/sites/default/files/publications/IPA%20Evidence%20Review_%20Remote%20Data%20Collection%20Modes.pdf.

Himelein, Kristen, Stephanie Eckman, Jonathan Kastelic, Kevin McGee, Michael Wild, Nobuo Yoshida, and Johannes Hoogeveen. 2020. "High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19. Guidelines on Sampling Design." Washington D.C.: World Bank. http://documents.worldbank.org/curated/en/742581588695955271/Guidelines-on-Sampling-Design.

Jolliffe, Dean, Daniel Gerszon Mahler, Malarvizhi Veerappan, Talip Kilic, and Philip Wollburg. 2023. "What Makes Public Sector Data Valuable for Development?" *The World Bank Research Observer*, April, lkad004. https://doi.org/10.1093/wbro/lkad004.

Kanyanda, Shelton, Yannick Markhof, Philip Wollburg, and Alberto Zezza. 2021. "Acceptance of COVID-19 Vaccines in Sub-Saharan Africa: Evidence from Six National Phone Surveys." *BMJ Open* 11 (12): e055159. https://doi.org/10.1136/bmjopen-2021-055159.

Keats, Anthony. 2018. "Women's Schooling, Fertility, and Child Health Outcomes: Evidence from Uganda's Free Primary Education Program." *Journal of Development Economics* 135 (November): 142–59. https://doi.org/10.1016/j.jdeveco.2018.07.002.

Kusuma, Dian, Hasbullah Thabrany, Budi Hidayat, Margaret McConnell, Peter Berman, and Jessica Cohen. 2017. "New Evidence on the Impact of Large-Scale Conditional Cash Transfers on Child Vaccination Rates: The Case of a Clustered-Randomized Trial in Indonesia." *World Development* 98 (October): 497–505. https://doi.org/10.1016/j.worlddev.2017.05.007.

Laajaj, Rachid, and Karen Macours. 2021. "Measuring Skills in Developing Countries." *Journal of Human Resources* 56 (4): 1254–95. https://doi.org/10.3368/jhr.56.4.1018-9805R1.

Lazarus, Jeffrey V., Scott C. Ratzan, Adam Palayew, Lawrence O. Gostin, Heidi J. Larson, Kenneth Rabin, Spencer Kimball, and Ayman El-Mohandes. 2021. "A Global Survey of Potential Acceptance of a COVID-19 Vaccine." *Nature Medicine* 27 (2): 225–28. https://doi.org/10.1038/s41591-020-1124-9.

Lazarus, Jeffrey V., Katarzyna Wyka, Trenton M. White, Camila A. Picchio, Lawrence O. Gostin, Heidi J. Larson, Kenneth Rabin, Scott C. Ratzan, Adeeba Kamarulzaman, and Ayman El-Mohandes. 2023. "A Survey of COVID-19 Vaccine Acceptance across 23 Countries in 2022." *Nature Medicine*, January. https://doi.org/10.1038/s41591-022-02185-4.

Lazarus, Jeffrey V., Katarzyna Wyka, Trenton M. White, Camila A. Picchio, Kenneth Rabin, Scott C. Ratzan, Jeanna Parsons Leigh, Jia Hu, and Ayman El-Mohandes. 2022. "Revisiting COVID-19 Vaccine Hesitancy around the World Using Data from 23 Countries in 2021." *Nature Communications* 13 (1): 3801. https://doi.org/10.1038/s41467-022-31441-x.

Levine, Gillian, Amadu Salifu, Issah Mohammed, and Günther Fink. 2021. "Mobile Nudges and Financial Incentives to Improve Coverage of Timely Neonatal Vaccination in Rural Areas (GEVaP Trial): A 3-Armed Cluster Randomized Controlled Trial in Northern Ghana." Edited by Patricia Evelyn Fast. *PLOS ONE* 16 (5): e0247485. https://doi.org/10.1371/journal.pone.0247485.

Li, Minghui, Ilene Harris, and Z. Kevin Lu. 2015. "Differences in Proxy-Reported and Patient-Reported Outcomes: Assessing Health and Functional Status among Medicare Beneficiaries." *BMC Medical Research Methodology* 15 (1): 62. https://doi.org/10.1186/s12874-015-0053-7.

Lim, Stephen S, David B Stein, Alexandra Charrow, and Christopher JL Murray. 2008. "Tracking Progress towards Universal Childhood Immunisation and the Impact of Global Initiatives: A Systematic Analysis of Three-Dose Diphtheria, Tetanus, and Pertussis Immunisation Coverage." *The Lancet* 372 (9655): 2031–46. https://doi.org/10.1016/S0140-6736(08)61869-3.

Mathieu, Edouard, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. 2021. "A Global Database of COVID-19 Vaccinations." *Nature Human Behaviour* 5 (7): 947–53. https://doi.org/10.1038/s41562-021-01122-8.

Miles, Melody, Tove K. Ryman, Vance Dietz, Elizabeth Zell, and Elizabeth T. Luman. 2013. "Validity of Vaccination Cards and Parental Recall to Estimate Vaccination Coverage: A Systematic Review of the Literature." *Vaccine* 31 (12): 1560–68. https://doi.org/10.1016/j.vaccine.2012.10.089.

Miller, Grant, and B. Piedad Urdinola. 2010. "Cyclicality, Mortality, and the Value of Time: The Case of Coffee Price Fluctuations and Child Survival in Colombia." *Journal of Political Economy* 118 (1): 113–55. https://doi.org/10.1086/651673.

Mosely, Ray R., and Fredric D. Wolinsky. 1986. "The Use of Proxies in Health Surveys: Substantive and Policy Implications." *Medical Care* 24 (6): 496–510.

Murray, Christopher JL, Bakhuti Shengelia, Neeru Gupta, Saba Moussavi, Ajay Tandon, and Michel Thieren. 2003. "Validity of Reported Vaccination Coverage in 45 Countries." *The Lancet* 362 (9389): 1022–27. https://doi.org/10.1016/S0140-6736(03)14411-X.

Nguyen, Kimberly H., Peng-Jun Lu, Seth Meador, Hung, Katherine Kahn, Jessica Hoehner, Hilda Razzaghi, Carla Black, and James A. Singleton. 2021. "Comparison of COVID-19 Vaccination Coverage Estimates from the Household Pulse Survey, Omnibus Panel Surveys, and COVID-19 Vaccine Administration Data, United States, March 2021." Centers for Disease Control and Prevention. https://www.cdc.gov/vaccines/imz-managers/coverage/adultvaxview/pubs-resources/covid19-coverage-estimates-comparison.html.

Palloni, Giordano. 2017. "Childhood Health and the Wantedness of Male and Female Children." *Journal of Development Economics* 126 (May): 19–32. https://doi.org/10.1016/j.jdeveco.2016.11.005.

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108 (11): 3266–3302. https://doi.org/10.1257/aer.20171330.

Reza, Agarwal, F. Sultana, R. Bari, and Ahmed Mushfiq Mobarak. 2022. "Why Vaccination Rates Are Lagging in Low- and Middle-Income Countries, and What Can We Do About It?" *BMJ*.

Reza, Hasan Mahmud, Vaishnavi Agarwal, Farhana Sultana, Razmin Bari, and Ahmed Mushfiq Mobarak. 2022. "Why Are Vaccination Rates Lower in Low and Middle Income Countries, and What Can We Do about It?" *BMJ*, July, e069506. https://doi.org/10.1136/bmj-2021-069506.

Rosenbaum, Rob, and Adele Waugaman. 2022. "Vaccine Administration Data Backlogs: Preliminary Findings & Next Steps." Presented at the COVID-19 Vaccine Delivery Partnership (CoVDP) meeting series, July 21. https://www.digitalhealthcoe.org/webinars-and-presentations/vaccine-administration-data-backlogs%3A-preliminary-findings-%26-next-steps.

Sandefur, Justin, and Amanda Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." *The Journal of Development Studies* 51 (2): 116–32. https://doi.org/10.1080/00220388.2014.968138.

Scobie, Heather M., Michael Edelstein, Edward Nicol, Ana Morice, Nargis Rahimi, Noni E. MacDonald, M. Carolina Danovaro-Holliday, and Jaleela Jawad. 2020. "Improving the Quality and Use of Immunization and Surveillance Data: Summary Report of the Working Group of the Strategic Advisory Group of Experts on Immunization." *Vaccine* 38 (46): 7183–97. https://doi.org/10.1016/j.vaccine.2020.09.017.

Shapira, Gil, Tashrik Ahmed, Salomé Henriette Paulette Drouard, Pablo Amor Fernandez, Eeshani Kandpal, Charles Nzelu, Chea Sanford Wesseh, et al. 2021. "Disruptions in Maternal and Child Health Service Utilization during COVID-19: Analysis from Eight Sub-Saharan African Countries." *Health Policy and Planning* 36 (7): 1140–51. https://doi.org/10.1093/heapol/czab064.

Solís Arce, Julio S., Shana S. Warren, Niccolò F. Meriggi, Alexandra Scacco, Nina McMurry, Maarten Voors, Georgiy Syunyaev, et al. 2021. "COVID-19 Vaccine Acceptance and Hesitancy in Low- and Middle-Income Countries." *Nature Medicine*, July. https://doi.org/10.1038/s41591-021-01454-y.

Stoop, Nik, Marijke Verpoorten, and Koen Deconinck. 2019. "Voodoo, Vaccines, and Bed Nets." *Economic Development and Cultural Change* 67 (3): 493–535. https://doi.org/10.1086/698308.

Struminskaya, Bella, and Michael Bosnjak. 2021. "Panel Conditioning: Types, Causes, and Empirical Evidence of What We Know So Far." In *Wiley Series in Probability and Statistics*, edited by Peter Lynn, 1st ed., 272–301. Wiley. https://doi.org/10.1002/9781119376965.ch12.

Triplett, Tim R. 2010. "Can Your Spouse Accurately Report Your Activities? An Examination of Proxy Reporting." *Survey Practice* 3 (1): 1–6. https://doi.org/10.29115/SP-2010-0002.

WHO. 2020a. "Immunization Agenda 2030: A Global Strategy to Leave No One Behind." Geneva: World Health Organization. https://cdn.who.int/media/docs/default-source/immunization/strategy/ia2030/ia2030-draft-4-wha_b8850379-1fce-4847-bfd1-5d2c9d9e32f8.pdf?sfvrsn=5389656e_69&download=true.

———. 2020b. "WHO COVID-19 Dashboard." World Health Organization. 2020. https://covid19.who.int/info.

Wollburg, Philip, Ivette Contreras, Calogero Carletto, Luis Gonzalez Morales, Francesca Perucci, and Alberto Zezza. 2022. "The Uneven Effects of the COVID-19 Pandemic on National Statistical Offices: Evidence from the Global COVID-19 Survey of NSOs." *Statistical Journal of the IAOS* 38 (3): 785–803. https://doi.org/10.3233/SJI-220044.

Wollburg, Philip, Yannick Markhof, Shelton Kanyanda, and Alberto Zezza. 2022. "Turning COVID-19 Vaccines into Vaccinations: New Evidence from Sub-Saharan Africa." Policy Research Working Paper Series 10152. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-10152.

———. 2023. "The Evolution of COVID-19 Vaccine Hesitancy in Sub-Saharan Africa : Evidence from Panel Survey Data." 10275. Policy Research Working Paper. Washington, D.C.: World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099622001122358675/idu040a460160167704ede0bf7b03ec8150698fb.

Wolter, Felix, Jochen Mayerl, Henrik K. Andersen, Theresa Wieland, and Justus Junkermann. 2022. "Overestimation of COVID-19 Vaccination Coverage in Population Surveys Due to Social Desirability Bias: Results of an Experimental Methods Study in Germany." *Socius: Sociological Research for a Dynamic World* 8 (January): 237802312210947. https://doi.org/10.1177/23780231221094749.

World Bank. n.d. "Statistical Performance Indicators (SPI): Overall Score (Scale 0-100)." World Bank Open Data. Accessed April 12, 2023. https://data.worldbank.org/indicator/IQ.SPI.OVRL.

Zezza, Alberto, Kevin Mcgee, Philip Wollburg, Thomas Assefa, and Sydney Gourlay. 2022. *From Necessity to Opportunity: Lessons for Integrating Phone and in-Person Data Collection for Agricultural Statistics in a Post-Pandemic World*. Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-10168.

Appendix

Table A1: Country coverage

Country	Number waves	Data collection period
Argentina	1	05/21-07/21
Belize	1	05/21-07/21
Bolivia	1	05/21-07/21
Brazil	1	07/21-09/21
Bulgaria	2	07/21; 10/21
Burkina Faso	3	04/22-05/22; 09/22; 12/22-01/23
Colombia	1	05/21-07/21
Costa Rica	1	05/21-07/21
Dominica	1	05/21-07/21
Dominican Republic	1	05/21-07/21
Ecuador	1	05/21-07/21
El Salvador	1	05/21-07/21
Ethiopia	1	12/23-01/22
Gambia, The	3	09/21; 10/21-11/21; 11/21-12/21
Guatemala	1	05/21-07/21
Guyana	1	05/21-07/21
Haiti	1	05/21-07/21
Honduras	1	05/21-07/21
Indonesia	2	03/21; 10/21
Iraq	3	06/21; 07/21; 09/21
Jamaica	1	05/21-07/21
Kenya	1	11/21-03/22
Lao PDR	1	11/21;
Malawi	3	04/21; 02/22; 07/22-09/22
Malaysia	2	05/21-06/21; 10/21
Mexico	1	05/21-07/21
Nicaragua	1	05/21-07/21
Nigeria	3	11/21-01/22; 03/22-04/22; 08/22
Paraguay	1	05/21-07/21
Peru	1	05/21-07/21
Philippines	1	05/21-06/21
St Lucia	1	05/21-07/21
Tanzania	1	12/21;
Thailand	1	05/21;
Uganda	2	09/21-11/21; 08/22
Uzbekistan	8	04/21; 05/21; 06/21; 09/21; 10/21; 11/21; 01/22; 02/22

Table A2: Proxy reporting accuracy

	Burkina Faso	Malawi	Uganda
Vaccination status	93.1	80.7	93.6
	(91.3 to 94.8)	(77.8 to 83.7)	(92.1 to 95.2)
Number of doses received	95.7	69.9	87.5
	(92.7 to 98.6)	(63.2 to 76.5)	(84.9 to 90.1)
Vaccine acceptance	53.2	55.6	54.1
	(48.9 to 57.5)	(40.5 to 70.7)	(48.3 to 59.8)
N (Vaccination status)	808	690	975
N (Number of doses)	185	186	633
N (Vaccine acceptance)	515	45	296

Note: Share of accurate proxy reports by main respondent about the random respondent when compared with self-reported information from the random respondent. Estimates exclude cases where the main respondent was identical to the random respondent. Accuracy for the number of doses received includes only cases where the main respondent correctly reported the vaccination status in the first place. Accuracy for vaccine acceptance only includes cases where the main respondent reported that the random respondent was not vaccinated. All values in percent. 95%-Confidence Intervals in parentheses.

Table A3: Summary of results

Issue	Sample	Survey Estimate	Admin Data
Survey estimates vs. administrative figures			
Alignment between survey and admin data	Purposively selected main respondents	7.7 pp difference (85.8%) 10.8 pp difference (98.9%) in absolute terms	--
Baseline Phone Survey Estimates	Purposively selected main respondents	BFA: 40.0% (05/22), 49.9% (09/22), 53.2% (01/23) ETH: 54.1 MWI: 9.7% (04/21), 39.2% (02/22), 51.6% (09/22) NGA: 26.5% (01/22), 49.7% (04/22), 53.7% (08/22) UGA: 34.9% (11/21), 86.0% (09/22)	BFA: 19.3% (05/22), 24.3% (09/22), 35.7% (02/23) ETH: 59.1% (01/22) MWI: 2.3% (04/21), 13.1% (02/22), 26.2% (09/22) NGA: 10.3% (01/22), 18.2% (04/22), 32.3% (08/22) UGA: 14.8% (11/21), 71.0% (09/22)
Survey Data: Errors of representation			
H1: Household Sample Selection	General population sample	29.5%	20.1%
H2: Respondent Selection	Randomly selected respondents	BFA: 34.8%, MWI: 45.2%, UGA: 78.0%	BFA: 24.3%, MWI: 26.2%, UGA: 71.0%
	All HH members (proxy-reported)	BFA: 19.3%, MWI: 20.3%, UGA: 41.0%	BFA: 13.6%, MWI: 14.9%, UGA: 38.9%
Survey Data: Errors of measurement			
H3: Survey Mode	Phone survey HHs interviewed in-person	30.5%	20.1%
H4: Panel Conditioning	First-time respondents	26.0%	10.3%
H5: Proxy Reporting Accuracy	All HH members (proxied but using random respondents' self-reports where possible)	BFA: 19.5%, MWI: 21.5%, UGA: 41.3%	BFA: 13.6%, MWI: 14.9%, UGA: 38.9%
H6: Experimenter Demand	Postive demand framing Negative demand framing	BFA: 50.1%, ETH: 49.0% BFA: 50.3%, ETH: 50.4%	BFA: 35.7%, ETH: 59.1%

Note: Summary of estimated vaccine coverage rates under different survey design choices and comparison to administrative figures. All estimates weighted with re-calibrated household weights (Ambel et al. 2021) except those under H6 where the phone survey sample was split into treatment groups.