

# PROGRAM TARGETING WITH MACHINE LEARNING AND MOBILE PHONE DATA

EMILY L. AIKEN, GUADALUPE BEDOYA, JOSHUA E. BLUMENSTOCK, AND AIDAN COVILLE



*This brief is based on Aiken, E. L., Bedoya, G., Blumenstock, J. E., and Coville, A. (2022). "Program Targeting with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan." Policy Research working paper; no. WPS 10252 Washington, D.C.: World Bank Group.*

## The Development Challenge

Each year, billions of dollars are spent on targeted social protection programs.<sup>1</sup> While these programs are often designed to provide support to the poorest households, identifying eligible households within the population poses a challenge.<sup>2</sup> The difficulty and cost of collecting data, and the variable quality of what gets collected, can introduce significant errors in the targeting process, and this is exacerbated in contexts of fragility, conflict, and violence (FCV), where data collection can be particularly challenging. This means that many ineligible households receive support (errors of inclusion), while eligible households are excluded (errors of exclusion). In this study, we explore whether machine learning applications using non-traditional mobile phone data can be leveraged as a complement to or a substitute for traditional targeting methods to improve the likelihood that programs reach their intended recipients.

## Study Design

We combine rich survey data from a "big push" anti-poverty program in Afghanistan with detailed mobile phone logs (call detail records, or CDR) from program recipients. First, we calculate several behavioral indicators of mobile phone use from the CDR. Then, we study the extent to which machine learning methods can accurately differentiate ultra-poor households eligible for program benefits from ineligible households.

Survey data collected in 2016 from 2,852 households include both ultra-poor (UP) and non-ultra-poor (non-UP) households as determined by the program's targeting. Of these, 535 households could be matched with CDR data from a large mobile operator in Afghanistan. From November 2015

to April 2016 (the period during which the targeting data and the household survey data were collected), these 535 households recorded 629,543 phone transactions (calls, texts, and phone recharges). From these data, 797 behavioral indicators were computed to measure communication patterns, social network structure, and spatial mobility. We then applied machine learning algorithms to these data to predict whether a household is ultra-poor on the basis of these behavioral indicators. The intuition is that ultra-poor individuals use their phones very differently from non-ultra-poor individuals and that machine learning algorithms can use those differences to predict ultra-poor status. Figure 1 presents an illustration of a CDR-based behavioral indicator: the number of days of active phone use by UP status, with a larger proportion of non-UP households reporting more active days.

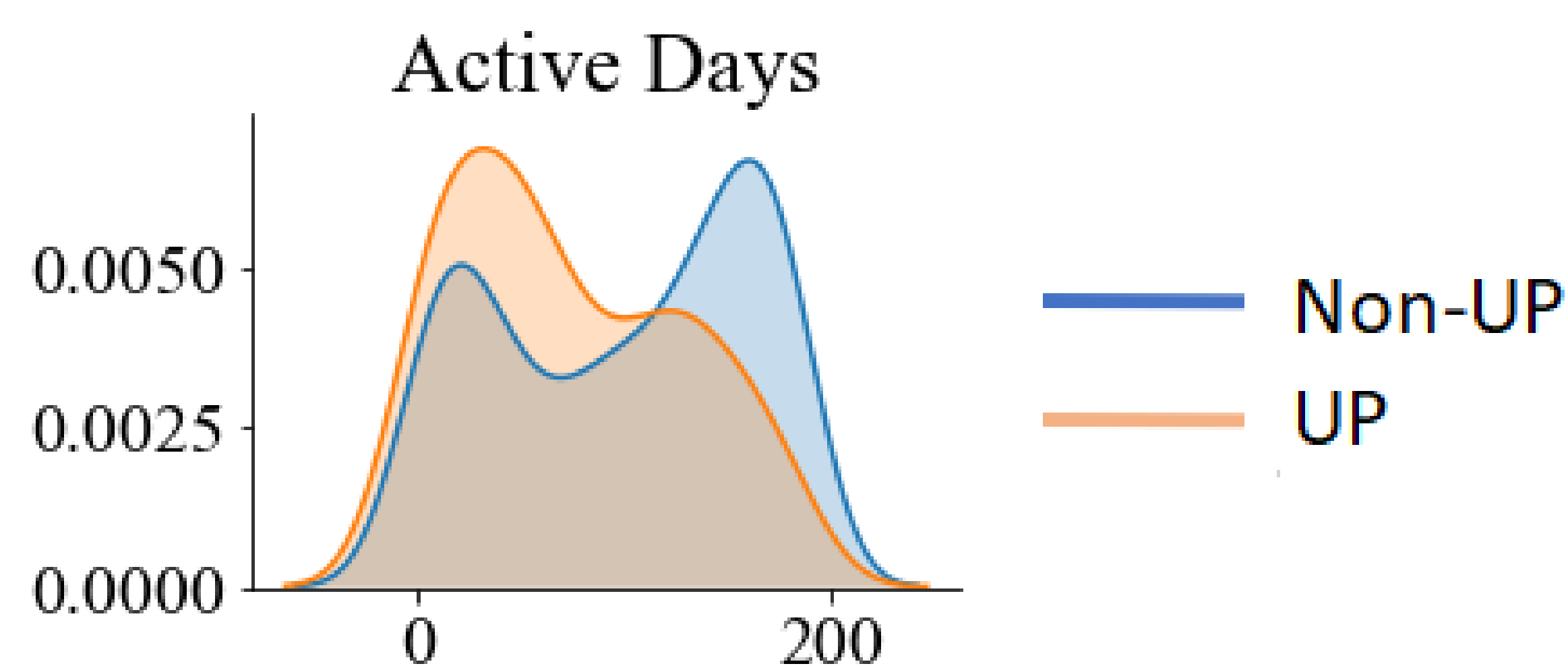


Figure 1: Example of a CDR-based behavioral indicator: active days density of UP vs. Non-UP households

The performance of this approach (*CDR-based method*) is then compared to the performance of more traditional targeting approaches using household assets (*asset-based wealth index*) or a measure of consumption consistent with what was used to determine official poverty levels in Afghanistan during the same period. The performance of these methods is assessed against the government's *hybrid targeting method*

<sup>1</sup> Gentilini, U., Almenfi, M., Orton, I., and Dale, P. (May 2020). "Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures." World Bank Policy Brief.

<sup>2</sup> Hanna, R., and Olken, B. (2018). "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries." *Journal of Economic Perspectives* 32, 201–226; and Lindert, K., Karippacheril, T. G., Caillava, I. R., and Chavez, K. N. (2020). *Sourcebook on the Foundations of Social Protection Delivery Systems*. World Bank Publications.

to identify ultra-poor households, consisting of a community-wealth ranking of the population in each village, plus a short in-person follow-up survey to verify vulnerability.

## Results

### *Machine learning methods leveraging mobile phone data are comparable to and cheaper than survey-based consumption and assets methods for identifying UP households.*

Figure 2 compares UP targeting performance by method. This indicator evaluates how accurate each method is for distinguishing ultra-poor from non-ultra-poor households versus the hybrid method used by the government. A value of 1.00 indicates perfect targeting, and 0.50 indicates targeting no better than random selection. Deviations from 1.00 indicate targeting errors.<sup>3</sup> Performance ranges from 0.68 for the CDR-based method to 0.73 for the asset-based index. Overall, we find that, while all methods (CDR, assets, and consumption) are imperfect, they are similarly effective at identifying UP households. Performance results are maintained if households without phones are classified as UP.<sup>4</sup>

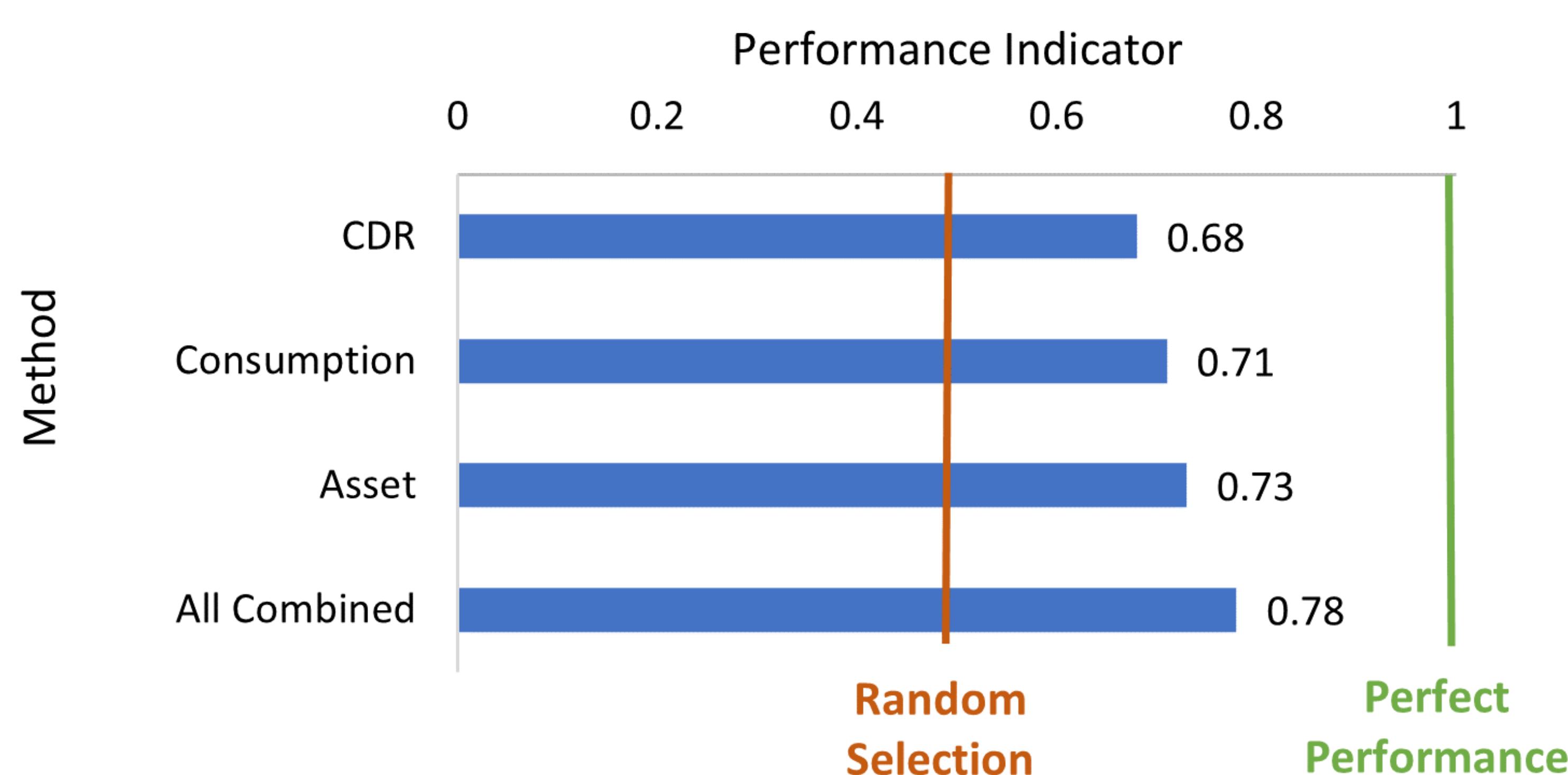


Figure 2: Targeting Performance by Method

### *Combining survey-based measures with mobile phone data produces classifications more accurate than those based on a single data source.*

The last bar in figure 2 represents the UP and non-UP classification when using all data sources together, showing that this approach is more accurate than any data source used on its own. This might be due in part to the fact that the hybrid method is more multidimensional, informed by community perceptions of vulnerability, while consumption and assets

methods are more limited, focusing on flows and stocks, respectively.<sup>5</sup> Also, it is worth noting that consumption-based targeting is rare in practice due to its high cost and logistical complexity for large populations. Short survey- and observation-based proxy-means measures are more commonly used. Furthermore, the feasibility of survey-based methods can be compromised in fragile contexts or during shocks that restrict mobility, such as the COVID-19 pandemic.

## Implications

These results suggest there is potential for using CDR-based methods to determine eligibility for development or humanitarian interventions, substantially reducing program targeting overhead and costs. Our results also indicate that CDR-based methods may complement and enhance existing survey-based methods, when they are feasible. However, a number of important factors need to be considered when assessing the appropriateness of integrating such a targeting approach.

- ▶ First, in settings where mobile phone penetration is limited, this approach is likely to induce errors if non-phone-owning households are excluded.
- ▶ Second, access to CDR data presents important privacy concerns; these can be addressed, but typically not without reducing the performance of such approaches.
- ▶ Third, to the extent that households preempt the targeting method and adapt their phone use accordingly, this can increase measurement error.

While these challenges are by no means unique to the use of CDR data, they do pose important practical constraints to implementing the targeting approach presented here. It is important to consider these limitations and the constraints of specific local contexts alongside the efficiency gains offered by CDR-based targeting.

## Acknowledgments

This work was supported by DARPA and the NIWC under contract N66001-15-C-4066, by the NSF under grant IIS-1942702, and by the World Bank's Knowledge for Change Program. The US government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense, the US government, or the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

<sup>3</sup> This performance measure is estimated as the area under the receiver operating characteristic curves for each method, which is a measure of targeting quality based on the trade-off between targeting errors (inclusion of non-UP and exclusion of UP) at different cutoff thresholds for UP.

<sup>4</sup> Phone penetration in this sample is 84 percent at the household level.

<sup>5</sup> Sen, A. (1992). "The Political Economy of Targeting." World Bank, Washington, DC; and Alkire, S., Foster, J., Seth, S., Santos, M. E., Roche, J. M., and Ballon, P. (2015). *Multidimensional Poverty Measurement and Analysis*. Oxford University Press.