DISCUSSION PAPER SERIES

DP16385

Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance

Emily Aiken, Suzanne Bellue, Christopher Udry, Dean Karlan and Joshua Blumenstock

DEVELOPMENT ECONOMICS

PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance

Emily Aiken, Suzanne Bellue, Christopher Udry, Dean Karlan and Joshua Blumenstock

Discussion Paper DP16385
Published 21 July 2021
Submitted 15 July 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance

Abstract

The COVID-19 pandemic has devastated many low- and middle-income countries (LMICs), causing widespread food insecurity and a sharp decline in living standards. In response to this crisis, governments and humanitarian organizations worldwide have mobilized targeted social assistance programs. Targeting is a central challenge in the administration of these programs: given available data, how does one rapidly identify the individuals and families with the greatest need? This challenge is particularly acute in the large number of LMICs that lack recent and comprehensive data on household income and wealth. Here we show that non-traditional “big” data from satellites and mobile phone networks can improve the targeting of anti-poverty programs. Our approach uses traditional survey-based measures of consumption and wealth to train machine learning algorithms that recognize patterns of poverty in non-traditional data; the trained algorithms are then used to prioritize aid to the poorest regions and mobile subscribers. We evaluate this approach by studying Novissi, Togo’s flagship emergency cash transfer program, which used these algorithms to determine eligibility for a rural assistance program that disbursed millions of dollars in COVID-19 relief aid. Our analysis compares outcomes – including exclusion errors, total social welfare, and measures of fairness – under different targeting regimes. Relative to the geographic targeting options considered by the Government of Togo at the time, the machine learning approach reduces errors of exclusion by 4-21%. Relative to methods that require a comprehensive social registry (a hypothetical exercise; no such registry exists in Togo), the machine learning approach increases exclusion errors by 9-35%. These results highlight the potential for new data sources to contribute to humanitarian response efforts, particularly in crisis settings when traditional data are missing or out of date.

JEL Classification: I32, I38, O12, O38, C55

Keywords: targeting, Machine Learning, poverty, mobile phone data

Emily Aiken - emilyaiken@berkeley.edu
University of California, Berkeley

Suzanne Bellue - sbellue@mail.uni-mannheim.de
University of Mannheim

Christopher Udry - christopher.udry@northwestern.edu
Northwestern University and CEPR

Dean Karlan - karlan@northwestern.edu
Northwestern University and CEPR

Joshua Blumenstock - jblumenstock@gmail.com
University of California, Berkeley and CEPR

Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance

Emily Aiken[1*], Suzanne Bellue[2], Dean Karlan[34], Chris Udry[3], Joshua Blumenstock[1*+]

Affiliations:

[1] University of California, Berkeley, CA 94720.

[2] University of Mannheim

[3] Northwestern University

[4] Innovations for Poverty Action

* These authors contributed equally to this work

[+] Correspondence to: jblumenstock@berkeley.edu

The COVID-19 pandemic has devastated many low- and middle-income countries (LMICs), causing widespread food insecurity and a sharp decline in living standards[1]. In response to this crisis, governments and humanitarian organizations worldwide have mobilized targeted social assistance programs[2]. *Targeting* is a central challenge in the administration of these programs: given available data, how does one rapidly identify the individuals and families with the greatest need[3]? This challenge is particularly acute in the large number of LMICs that lack recent and comprehensive data on household income and wealth[4–6]. Here we show that non-traditional "big" data from satellites and mobile phone networks can improve the targeting of anti-poverty programs. Our approach uses traditional survey-based measures of consumption and wealth to train machine learning algorithms that recognize patterns of poverty in non-traditional data; the trained algorithms are then used to prioritize aid to the poorest regions and mobile subscribers. We evaluate this approach by studying Novissi, Togo's flagship emergency cash transfer program, which used these algorithms to determine eligibility for a rural assistance program that disbursed millions of dollars in COVID-19 relief aid. Our analysis compares outcomes – including exclusion errors, total social welfare, and measures of fairness – under different targeting regimes. Relative to the geographic targeting options considered by the Government of Togo at the time, the machine learning approach reduces errors of exclusion by 4-21%. Relative to methods that require a comprehensive social registry (a hypothetical exercise; no such registry exists in Togo), the machine learning approach increases exclusion errors by 9-35%. These results highlight the potential for new data sources to contribute to humanitarian response efforts, particularly in crisis settings when traditional data are missing or out of date.

The COVID-19 pandemic has led to a sharp decline in living standards across the world, as policies designed to stop the spread of the disease have disrupted ordinary economic activity. Economically vulnerable households in low- and middle-income countries (LMICs) are among the hardest hit, with over 100 million individuals estimated to have transitioned into extreme poverty since the onset of the pandemic[7].

To offset the most severe consequences of this sudden income decline, governments and humanitarian organizations around the world have mobilized relief efforts. Gentilini et al. (2021) estimates that over 3,300 new social assistance programs have been launched since early 2020[2], providing over $800 billion dollars in cash transfer payments to over 1.5 billion people (roughly one fifth of the world's population).

The overwhelming majority of COVID-19 response efforts – and the majority of cash transfer programs globally – provide *targeted* social assistance. In other words, specific criteria are used to determine potential eligibility, typically some proxy for socioeconomic status. In most wealthy nations, governments rely on recent household income data to determine program eligibility[8]. However, in low- and middle-income countries (LMICs), where economic activity is often informal and based on home-produced agriculture, governments typically do not observe income for the vast majority of the population[3]. Other potential sources of targeting data are often incomplete or out of date[6,9]; for example, only half of the poorest countries having completed a census in the past 10 years[4]. In such contexts, data gaps preclude governments from implementing well-targeted social assistance programs[5,10].

Here we develop, implement, and evaluate a new approach to targeting social assistance based on machine learning algorithms and non-traditional "big data" from satellites and mobile phone networks. This approach leverages recent advances in machine learning that show that such data can help accurately estimate the wealth of small geographic regions[4,11,12] and individual mobile subscribers[13,14]. It also builds on a rich economics literature on the design of appropriate mechanisms for targeting social assistance[3,15–18]. See Methods §1 for a discussion of prior work.

Our results are based on the design and evaluation of Novissi, Togo's flagship emergency social assistance program. The Government of Togo launched Novissi in April 2020, shortly after the first COVID-19 cases appeared in the country. As economic lockdown orders forced many Togolese to stop working and led to widespread food insecurity (Figure S1), Novissi aimed to provide subsistence cash relief to those most impacted.[i] Eligible beneficiaries received bi-weekly payments of roughly USD $10. In an effort to minimize in-person contact, Novissi enrollment and payment was digital: beneficiaries registered using their mobile phones and transfers were made via mobile money. Full details on the Novissi program are provided in Methods §2.

When the government first launched Novissi, it did not have a traditional social registry that could be used to assess program eligibility, and had neither the time nor the resources to build

---

[i] See https://novissi.gouv.tg/

such a registry in the middle of the pandemic.[ii] Instead, Novissi eligibility was determined based on data contained in a national voter registry that had been updated in late 2019. Specifically, benefits were initially disbursed to individuals who met three criteria: (1) "self-targeted"[15] by dialing in to the Novissi platform and entering basic information from their mobile phone and; (2) registered to vote in specific regions (the program initially focused on the Greater Lomé region around the capital city); and (3) self-declared to work in an informal occupation in their voter registration.[iii]

Our research efforts focused on helping the government expand the Novissi program from informal workers in Greater Lomé to poorer individuals in rural regions of the country, and were designed to meet the government's two stated policy objectives: first, to direct benefits to the poorest geographic regions of the country; and second, to prioritize benefits to the poorest mobile subscribers in those regions.[iv] The approach we developed, which uses machine learning to analyze non-traditional data from satellites and mobile phone networks, has two distinct steps (Figure 1).

In the first step, we constructed micro-gridded estimates of the average consumption of every 1km^2 region in Togo by applying machine learning algorithms to high-resolution satellite imagery. Our approach uses pre-trained computer vision algorithms to extract features from the satellite images, then trains a machine learning algorithm to predict the average consumption of households in those images, using nationally-representative household survey data as training data (Methods §3.b)[4,12].

In the second step, we estimated the average daily consumption of each mobile phone subscriber by applying machine learning algorithms to mobile phone metadata provided by Togo's two mobile phone operators (see Methods §3d for a discussion of data privacy concerns). Specifically, we conducted surveys with a large and representative sample of mobile phone subscribers, used the surveys to measure the wealth and/or consumption of each subscriber, and then matched the survey-based estimates to detailed metadata on each subscriber's history of phone use. This sample was used to train supervised machine learning algorithms that predict socioeconomic status from phone use (Pearson ρ ranges from 0.41-0.46 – see Methods §4)[13,14,20]. This second step is similar in spirit to a traditional proxy means test (PMT), with two main differences: we used a high-dimensional vector of mobile phone features instead of a low-dimensional vector of assets to estimate wealth; and we used machine learning algorithms

---

[ii] The most recent census, which was completed in 2011, did not contain information on household wealth or poverty; more recent national surveys on living standards only contacted a small fraction of all households. As discussed in Methods §1, social registries used to determine potential eligibility for anti-poverty programs in LMICs frequently rely on proxy means tests or related methods, where the eligibility of each household is determined by asking household members a number of questions about the assets and characteristics of the household. To construct a universal registry usually requires that government surveyors visit all households in the country, and typically takes years to organize and execute[19].

[iii] The decision to target informal occupations helped prioritize benefits to people who were forced to stop working at the onset of the crisis. However, this approach does not necessarily target benefits to the poorest households in the country (Figure S2 and Methods §5).

[iv] Individuals without access to a mobile phone could not receive Novissi payments, which were digitally delivered using mobile money. We discuss issues regarding exclusion of those without phones in Methods §5.f.

designed to maximize out-of-sample predictive power instead of the traditional linear regression that maximizes in-sample goodness-of-fit[21].

Our main analysis evaluates the performance of this new approach to targeting that combines machine learning and mobile phone data – which we refer to more succinctly as the *phone-based approach* – by comparing targeting errors using the phone-based approach to targeting errors under three counterfactual approaches: a geographic targeting approach that the government piloted in summer 2020 (where all individuals are eligible within the poorest *prefectures*, Togo's admin-2 level, or poorest *cantons*, Togo's admin-3 level); occupation-based targeting (including Novissi's original approach to targeting informal workers as well as an "optimal" approach to targeting the poorest occupation categories in the country); and a parsimonious method based on phone data without machine learning (that uses total expenditures on calling and texting as a proxy for wealth).

We present results that compare the effectiveness of these different targeting mechanisms in two different scenarios. First, we evaluate the actual policy scenario faced by the of Togo in September of 2020, which involved distributing cash to 60,000 beneficiaries within Togo's 100 poorest cantons (the smallest administrative unit in the country, each containing on average 10 villages). This first scenario is evaluated using data collected in a large phone survey we designed for this purpose and conducted in September 2020. The "ground truth" measure of poverty in this first scenario is a proxy-means test, as consumption data could not be feasibly collected in the phone survey.[v] Second, we simulate and evaluate a more general and hypothetical policy scenario in which the government is interested in targeting the poorest individuals nationwide; this scenario is evaluated using national household survey data collected in the field by the government in 2018-2019. The second simulation uses consumption as the "ground truth" measure of poverty. These data are described in Methods §3 and details on the evaluation are in Methods §5.

In the first scenario focused on reaching the poorest people in the 100 poorest cantons, we find that the phone-based approach to targeting significantly reduces errors of exclusion (true poor who are mistakenly deemed ineligible) and errors of inclusion (non-poor who are mistakenly deemed eligible), relative to the other feasible approaches to targeting available to the Government of Togo (Figure 2a and first four columns of Table 1). We focus on the ability of each targeting method to reach the poorest 29% in each of the two survey datasets, since the rural Novissi expansion only had sufficient funding to provide benefits to 29% of individuals in eligible geographies (Table S5 and Table S6 evaluate performance using alternative poverty thresholds). Using a PMT as a measure of "true" poverty status, phone-based targeting (AUC= 0.70) outperforms the other feasible methods of targeting rural Novissi aid (e.g., AUC=0.59-0.64

---

[v] The PMT is based on a stepwise regression procedure, described in Appendix A: Selection of Variables for Proxy-Means Test, which captures roughly 48% of the variation in consumption. Thus, for the first scenario focused on the rural Novissi program, all targeting methods are evaluated for targeting quality with respect to this PMT. The phone-based machine learning model is likewise trained using the PMT as ground truth.

for geographic blanket targeting). As a result, errors of exclusion (defined as 1 – Recall) are lower for the phone-based approach (53%) than for feasible alternatives (59%-81%).

Phone-based targeting likewise outperforms most feasible methods when we simulate the targeting of a hypothetical national anti-poverty program (Figure 2b and last four columns of Table 1). Here, the phone-based approach is more effective at prioritizing the poor (AUC = 0.73) than geography-based alternatives (AUC = 0.65 - 0.68), and likewise leads to lower exclusion errors (50%) than most feasible alternatives (52%-76%). One exception in this hypothetical program is occupation-based targeting: while the Novissi program's original criteria of targeting informal workers would not scale well to a national program (76% exclusion errors), an alternative "optimal" occupation-based approach that we develop (Methods §5.b), which assigns all transfers the poorest occupational category (agricultural workers), slightly outperforms phone-based targeting (48% exclusion errors).

Taken together, the results in Table 1 indicate that the phone-based targeting approach was more effective in the actual rural Novissi program than it would be in a hypothetical nationwide program. While there are several differences between the two sets of results that might explain this gap,[vi] our analysis suggests that the benefits of phone-based targeting are greatest when the population under consideration is more homogeneous, and when there is less variation in other factors (such as place of residence) that are used in more traditional approaches to targeting. For instance, when we restrict the simulation of the hypothetical national program to households in rural areas, the gains from phone-based targeting increase (Table S7).

We likewise find that the performance benefits of phone-based targeting increase as programs seek to target the most extreme poor. This can be seen by comparing Table 1, where targeted performance is measured by how many of the poorest 29% receive benefits, to Table S5, which measures whether households below the poverty line ($1.90 per capita daily consumption) are targeted, and Table S6, which focuses on households below the extreme poverty line ($1.35 per capita daily consumption). While all targeting methods perform better at targeting the extreme poor, the differential between the phone-based approach and other methods is greater when the consumption threshold is lower.

The phone-based approach we develop relies heavily on machine learning to construct a poverty score for each mobile subscriber, where eligibility is a complex function of how the subscriber uses their phone (Figure S5). We also consider an alternative approach that does not use machine learning, but instead simply targets mobile phone subscribers with the lowest mobile phone expenditures over the preceding months (Methods §4.b). We find that this "phone expenditure" model (AUC= 0.57 for rural Novissi and 0.63 in for the hypothetical national program – see Table 1) performs substantially worse than the ML-based model (AUC= 0.70 for rural Novissi

---

[vi] These differences include: the rural Novissi evaluation is focused in 100 cantons while the hypothetical national program is evaluated nationwide; the rural Novissi evaluation uses a PMT as ground truth while the national evaluation a consumption measure; the rural Novissi evaluation is representative of mobile network subscribers while the national evaluation is representative of the entire population; the rural Novissi evaluation uses data from 2020 while the national evaluation's data is from 2018-2019; and the phone survey used to collect data for the rural Novissi evaluation is likely of a lower quality than the in-person survey used for the national evaluation.

and 0.73 in for the hypothetical national program). While the phone expenditure model requires much less data and may be easier to implement, this parsimony increases targeting errors – and may also introduce scope for strategic "gaming" if used repeatedly over time.

We also compare the phone-based approach to alternative targeting approaches that require a recent and comprehensive social registry. While the Government of Togo did not have such a registry, this comparison helps situate this method relative to other methods commonly used by development researchers and policymakers. These results, shown in Panel B of Table 1, can only be simulated using the national in-person survey, since the phone survey did not collect consumption data). The results are more ambiguous: the phone-based approach (AUC = 0.70-0.73) is approximately as accurate as targeting using an *asset-based wealth index* (AUC = 0.54-0.75), but less accurate than using a *poverty probability index* (AUC = 0.81) or perfectly-calibrated *proxy-means test* (AUC = 0.85).[vii] We note, however, that the performance of the "perfectly calibrated" PMT may substantially over-estimate the performance of a real-world PMT, which declines steadily over time since calibration (Methods §5.b)[22,23].

Improvements in targeting performance translate to an increase in social welfare. Using the constant relative risk-aversion (CRRA) utility function, we calculate aggregate welfare under the phone-based approach and each of the counterfactual targeting approaches. Under the CRRA assumptions, individual utility is a concave function of consumption. By assuming a fixed budget – which we fix at a size analogous to that of the Novissi rural aid program, which had a budget of USD 4 million to distribute among 154,238 program registrants – and equal transfer sizes to all beneficiaries, we simulate the distribution of benefits among eligible individuals at counterfactual targeting thresholds to trace out social welfare curves for each targeting method. This social welfare analysis also allows us to identify the optimal beneficiary share and corresponding transfer size. Figure 3 shows the utility curves for each of the targeting methods simulated, separately for the two datasets. Note that phone-based targeting, geographic blanketing, and an asset-based wealth index all achieve approximately the same maximum utility in the hypothetical national program, but phone-based targeting dominates in the rural Novissi program. Also note that all targeting methods outperform a universal basic income scheme if the beneficiary share and transfer size is well-calibrated.

These utilitarian welfare gains suggest that society as a whole will benefit from improved targeting, but do not imply that all subgroups of the population will benefit equally. Indeed, there is growing concern that algorithmic decision-making can unfairly discriminate against vulnerable groups[24–26]. To address these concerns in the context of the Novissi program, we audit the fairness of each targeting method across a set of potentially sensitive characteristics, while noting that notions of fairness and parity are contested and often in tension.[27] Figure 4a shows, as an example, that the phone-based approach does not cause women to be systematically more likely to be incorrectly excluded by the targeting mechanism from receiving benefits than men (see also Methods §5.e). Likewise, the phone-based approach does not create significant exclusion errors

---

[vii] The asset-based wealth index uses principal component analysis to generate an index of poverty based on ownership of 8-20 assets. The PPI is implemented using IPA's standard poverty scorecard for Togo (https://www.povertyindex.org/country/togo), based on point scoring for ten variables relating to poverty. The PMT uses machine learning to select eleven variables that are jointly most predictive of consumption. See Methods § 3.a.

for specific ethnic groups (Figure 4b), religions, age groups, or type of household, though there are small differences in targeting accuracy between groups (Figure S7). We also compare the fairness of the phone-based approach to several other targeting approaches by evaluating each method's demographic parity, i.e., the extent to which each method under- or over-targets specific demographic subgroups, relative to that group's true poverty rate (Figure 4c-d and Figure S8). Overall, we find that none of the targeting methods analyzed naively achieves perfect parity across subgroups; a phenomenon referred to as "no fairness through unawareness."[28] The largest parity differences occur with geographic targeting methods.

This novel approach to targeting requires careful consideration of the ways in which individuals can be incorrectly excluded from receiving program benefits (Methods §5.f). Our analysis highlights five main sources of exclusion errors for the expansion of Novissi (Table 2): (*i*) beneficiaries must have a SIM card and access to a mobile phone (2018-2019 field survey data indicate that 65% of adults and 85% of households have a phone; see also Figure S9); (*ii*) they must be a registered voter (roughly 87% of adults); (*iii*) they must self-target and attempt to register (roughly 40% of eligible individuals attempted); (*iv*) they must succeed in registering, which requires basic reading and digital literacy (72% succeed); (*v*) they must be successfully identified as eligible by the ML algorithm (47% recall, per Table 1). Many of these sources of possible exclusion overlap; Table S8 thus estimates, based on the 2020 phone survey, the extent to which each successive step in registration creates additional exclusions. These results highlight the fact that algorithmic targeting errors are an important source of program exclusion, but that real-world programs also face structural and environmental constraints to inclusion.

More broadly, our analysis shows how non-traditional "big" data can help identify the poor who are meant to benefit from an emergency social assistance program. The identification is imperfect: there are important errors of both exclusion and inclusion when relying on these data sources. There are also important limitations to this approach, for instance regarding data access and privacy; several such considerations are discussed in Methods §6. Moreover, our results do not imply that mobile phone-based targeting should replace traditional approaches reliant on proxy means tests or community-based targeting. Rather, these new methods provide a rapid and cost-effective supplement that may be most useful in crisis settings or in contexts where traditional data sources are incomplete or out of date. We believe future work should explore how real-time data sources, such as the phone data used by Novissi, can be best combined with more traditional field-based measurements, so that these complementary data sources can be best integrated in the design of inclusive systems for social protection.

Methods

1. Related work

There is a rich history of theoretical and empirical work that compares and evaluates methods for targeting social transfer programs. While there is increasing interest in "universal basic income", in which everyone is eligible for transfers, most countries use one or more targeting mechanisms to determine eligibility[3]. Typically, the goal of targeting is to ensure that the poorest individuals receive transfers.

Many programs include some degree of *self-targeting*, in which beneficiaries are required to take some action in order to receive benefits[15,29,30]. If the benefits of the program, relative to the costs associated with that action, are higher for poorer people, self-targeting can direct a greater share of benefits to the poor. *Geographic targeting* is also common, whereby benefits are restricted to individuals who live in specific regions[31,32]. Empirical evidence on geographic targeting indicates that more granularly targeted programs can be more effective at prioritizing the poor, but the implementation of such programs requires fine-grained poverty maps and distribution mechanisms that can be deployed in small regions[33–35]. With *proxy means tests (PMT)*, a number of variables are collected for each household, which are then used to impute an approximate measure of consumption or wealth for that household.[16,36] Likewise, a simple poverty scorecard or *poverty probability index (PPI)* uses a small number of variables to impute a poverty score.[22,37] PMTs and PPIs are frequently used in LMICs, but do require that the government collect and maintain a comprehensive social registry that records the information of each household. Finally, *community-based targeting (CBT)* approaches rely on members of the community to identify the poorest households in the area[38,39]. CBT-based approaches do not always target the lowest-consumption households, but allow the community to define their own notion of poverty, which can lead to higher satisfaction among community members[29] but may also lower perceptions of program legitimacy[40].

2. The COVID-19 pandemic in Togo and the Novissi program

Togo is a small country of roughly 8 million in West Africa. Over 50% of the population lives below the international poverty line. Shortly after the first COVID-19 cases were confirmed in Togo in early March 2020, the government imposed economic lockdown orders to prevent the spread of the disease. These lockdowns forced many Togolese to stop working, raising concerns about the potential for rising food insecurity (Figure S1).

On April 8 2020, the government launched the Novissi program, where "Novissi" means "solidarity" in the Ewé language. According to Minister Cina Lawson, Novissi "was built and designed in order to help those people who are the most vulnerable population and the most impacted by the anti-COVID measures."[viii] Novissi was initially designed to provide benefits to informal workers in Greater Lomé, the large metropolitan area surrounding the capital city where

[viii] https://undp-ric.medium.com/cina-lawson-a-covid-cash-transfer-programme-that-gives-more-money-to-women-in-togo-2386c5dff49

the lockdown orders were initially focused. The rationale for targeting informal workers was that they were more likely to be vulnerable, and more likely to be impacted by the lockdown orders.

To determine eligibility for Novissi, the government relied upon a national voter registry that was updated in late 2019, in which individuals indicated their home location and occupation. At the time, the voter registry contained 3,633,898 entries, which the electoral commission reports is equivalent to 87% of the total adult population (see Table 2 for details).

Receiving Novissi benefits required that individuals register by dialing in to the Novissi USSD platform from a mobile phone. Thus, registration initially required (i) a valid and unique voter ID linked to an eligible occupation from an eligible location; (ii) a valid SIM card, and (iii) access to a mobile phone. A smartphone was not required for registration; the USSD platform was accessible from a basic phone. Since phone sharing is common in Togo, multiple SIM cards could be registered through a single phone (so long as each SIM was then linked to a valid voter ID). See Methods §5.f for a discussion of the extent to which voter and phone requirement may have led to program exclusions.

Eligible female beneficiaries were then paid 12,250 FCFA (USD $22.50) per month; men received 10,500 FCFA (USD $20) per month. The payments were disbursed in two bi-weekly installments, for three months, using existing mobile money infrastructure managed by the country's two mobile network operators. The system was designed to be 100% digital, so that registration, eligibility determination, and payment could all be accomplished without face-to-face contact. Novissi was promoted actively through radio advertisements and community leaders, and 4.4 million registration attempts were reported on the day the program launched.[ix] In this first phase of Novissi, which focused on Greater Lomé, 511,611 beneficiaries received payments.

Our analysis focuses on a second phase of Novissi, which was initiated after the Novissi program in Greater Lomé had terminated. Specifically, in partnership with the NGO GiveDirectly, the government wished to expand Novissi eligibility to the rural poor. The policy mandate from the government was to (i) prioritize benefits to people living in Togo's 100 poorest cantons (of the 397 cantons nationally), where the number 100 was selected by the government in order to balance the desire to focus on the poorest villages, without focusing excessively on specific regions; and (ii) prioritize the poorest individuals in those 100 cantons.

During the second phase of Novissi, registration and enrollment used several of the same steps described above: individuals were required to have a voter ID registered in one of the 100 poorest cantons, and they had to self-register using a mobile phone with a unique SIM card. However, the individual's occupation was not used to determine eligibility; instead, the estimated wealth of the individual, based on the ML methods described in this paper, were used to limit eligibility to the estimated poorest subscribers in those 100 cantons.

[ix] https://undp-ric.medium.com/cina-lawson-a-covid-cash-transfer-programme-that-gives-more-money-to-women-in-togo-2386c5dff49

3. Data Sources

a. Survey Data

Our core analysis relies heavily on two surveys conducted by Togo's Institut National de la Statistique et des Etudes Economiques et Démographiques (INSEED). The first survey, which is nationally representative, was conducted in the field in 2018 and 2019 (N = 6,171). The second survey was conducted over the phone in September 2020, and is representative of mobile network subscribers inferred to be living in rural cantons eligible for Novissi aid (N = 8,915).[x]

2018-2019 Field Survey: Our first survey dataset was obtained from a nationally representative household survey. Specifically, 540 enumeration areas (EAs) were drawn at random from Togo's approximately 6,000 EAs, with weight proportional to the size of the EA in the last national census (conducted in 2011). 12 households were then drawn at random from each of the selected EAs to be interviewed, for a total of 6,172 households. Surveys, which lasted about three hours, were conducted in two waves, with the first wave between October and December 2018 and the second wave between April and June 2019. We remove one observation that is missing consumption expenditure and asset data, leaving 6,171 observations. Interviews took place with the head of household when possible, and alternatively with the most knowledgeable adult present.

As part of the survey's recontact protocol, phone numbers were requested from a representative of each household; 4,618 households (75%) of households are matched to a phone number. The data do not include an identifier for which member of the household the phone number belongs to. 4,171 households have phone numbers that contain at least one transaction in our mobile phone transaction logs in the three months prior to their survey date (91% of households with phone numbers), leading to a matched survey-mobile phone dataset of N = 4,171. Note that this matched dataset is not nationally representative nor necessarily representative of mobile phone subscribers, as there is selection in which households and household members provide phone numbers.

2020 Phone Survey: Our second survey dataset is obtained from a phone survey conducted over two weeks in September 2020. The survey lasted approximately 40 minutes, and covered demographics, asset ownership, and well-being. Phone numbers for the 2020 phone survey were drawn from mobile phone transaction logs and the sample is representative of subscribers inferred based on their mobile phone data to be living in rural cantons eligible for Novissi aid (See Appendix B: Sampling for 2020 Phone Survey). Note that since the sample is drawn based on inferred location, not all interviewees necessarily reside in an aid-eligible canton. The survey

[x] We use these two different survey datasets because neither dataset is sufficient by itself for the anlaysis we require: The 2020 survey did not collect consumption data, which is important for evaluating certain counterfactuals, and the 2018-19 survey only surveyed a small number of households in the 100 poorest cantons that were eligible for Novissi. We had planned to conduct a large in-person survey in early 2021 that would provide the single point of focus for this paper, but were forced to postpone the survey indefinitely due to a resurgence in COVID-19.

includes a question on canton of residence, and 68% of observations report living in a Novissi-eligible canton.

Of the phone numbers drawn, 28% respond, consent to the survey, and complete the entire survey. In total, after removing low-quality surveys and those missing poverty outcomes, the dataset contains 8,915 observations corresponding to individual subscribers. We reweight the survey for nonresponse using the same mobile phone features and machine learning methods described in Methods §4. Specifically, we train a gradient boosting model (implemented with LightGBM) to predict response from mobile phone features, and produce predictions of response probability out-of-sample over 5-fold cross validation. Our sample weights consist of the inverse of the draw probability and the inverse of the predicted probability of response.

Construction of Poverty Outcomes: We construct four poverty outcomes from the survey data: consumption expenditure (captured in the 2018-2019 field survey only), an asset-based wealth index, a poverty probability index (PPI), and a proxy-means test (PMT).

- *Consumption expenditure:* The consumption expenditure outcome is only available in the dataset from the 2018-2019 field survey. Disaggregated expenditures for more than 200 food and nonfood items are elicited in each household interview. The consumption aggregate is then adjusted for a price index calculated at the prefecture level. The final outcome measure is per-capita adult equivalent household consumption expenditure, which we transform to USD per day.
- *Asset index:* We calculate a PCA asset index for households in the 2018-2019 field survey and for the households associated with individuals interviewed in the 2020 phone survey. Asset indices are constructed with Principal Component Analysis (PCA). The asset index is constructed from 24 underlying binary asset variables in the 2018-2019 field survey and 10 underlying binary asset variables in the 2020 phone survey. The asset indices for the two surveys are constructed independently, from different sets of assets, and therefore do not share a basis vector. The basis vector for each index is shown in Table S1. The asset index explains 31.50% of the variance in asset ownership in the 2018-2019 field survey, and 53.45% of the variance in asset ownership in the 2020 phone survey. However, the variance explained in the two indices should not be directly compared since there are far fewer assets recorded in the 2020 phone survey than in the 2018-2019 field survey. We also note that the asset index for the 2020 phone survey dataset is dominated by variation in ownership of three assets (toilet, radio, and motorcycle; see Table S1) and is therefore considerably less smooth than the asset index in the 2018-2019 phone survey dataset.
- *Poverty probability index (PPI):* We use the scorecard for the current poverty probability index used by Innovations for Poverty Action (IPA).[xi] The index is calibrated based on a nationally representative survey conducted by INSEED in 2015 (N=2,335). "Poverty probability" is scored based on ten household questions, including region of residence, education of adults and children, asset ownership, and consumption of sugar. We

[xi] https://www.povertyindex.org/country/togo

calculate the PPI only for households in the 2018-2019 field survey, as the data necessary for all components were not collected in the 2020 phone survey.

- *Proxy-means test (PMT):* Using the data from the 2018-2019 field survey, we follow a stepwise forward selection process to select the 12 asset and demographic variables that are jointly most predictive of per-capita household consumption (see Appendix A: Selection of Variables for Proxy-Means Test for details). We use these variables to construct a consistent proxy-means test (PMT) for the 2018-2019 field survey and the 2020 phone survey. Following recent literature, we use a regularized linear model (Ridge regression) rather than a simple linear regression to maximize out-of-sample accuracy[21,26]. For the 2018-2019 field survey, PMT "consumption" estimates are produced out-of-sample over 10-fold cross validation. For the 2020 phone survey, we train the Ridge regression on the entire 2018-2019 field survey sample and use the fitted model to produce PMT "consumption" estimates for each phone survey observation. Over 10-fold cross validation, the PMT explains 48.35% of the variance in log-transformed consumption expenditure in the 2018-2019 field survey. The weights for the PMT are included in Table S2. Since they are trained to predict consumption, PMT "consumption" estimates can be interpreted as estimated USD/day.

Construction of Occupation Categories: We use self-reported occupation (of the household head for the 2018-2019 field survey, and of the respondent for the 2020 phone survey) to categorize occupations and later simulate occupation-based targeting. We first classify each of the self-reported occupations according to the occupation categories in the Novissi registry. We identify which of these categories are informal (in the Novissi registry, more than 2,000 unique occupations are considered informal – some of the most common ones are venders, hairdressers, taxi drivers, tailors, construction workers, and the unemployed). We further classify occupations in 10 broad categories according to the Afrostat system[xii]. Table S3 records these categories, along with the proportion in each category in each of the two surveys and associated poverty.

b. Poverty Maps

To simulate geographic targeting, we rely on poverty maps of Togo's prefectures (admin-2 level, 40 prefectures) and cantons (admin-3 level, 397 cantons). In the 2018-2019 field survey, the latitude and longitude of each household were recorded by enumerators as part of the interview, so we map each observation to a prefecture and canton using the geographic coordinates. For the 2020 phone survey, we ask each respondent to report their prefecture and canton of residence.

Prefecture poverty map: INSEED completed a survey-based poverty mapping exercise in 2017. Specifically, a proxy-means test was calibrated on a small consumption sample survey conducted in 2015 (N = 2,335). 26,902 households were then surveyed in the field over three weeks in 530 EAs, sampled to be representative at the prefecture level. The interview included questions on demographics, education, asset ownership, and household characteristics that made up the PMT. The calibrated PMT was then used to infer the "consumption" of each household, and observations were aggregated to estimate the percentage of the population living under the Togo-

[xii] https://www.afristat.org/nomenclatures/

specific poverty line of USD 1.79/day in each prefecture. Figure S3 shows the resulting poverty map. For validation, we evaluate the correlation between prefecture-level poverty rates from the poverty mapping exercise and average consumption in the 2018-2019 field survey. The Person correlation coefficient is -0.78, and the Spearman correlation coefficient is -0.70.

Canton poverty map: When COVID-19 first appeared in Togo in early 2020, it had been at least ten years since a household survey had been conducted in Togo that was representative at the canton level. Togo's last census was conducted in 2011, but did not include information on income, consumption, or asset ownership. We therefore rely on recently-produced publicly available satellite-based estimates of poverty which use deep learning models trained on DHS data from neighboring countries to estimate average wealth of each 1km^2 satellite tile in Togo.[12] We overlay the resulting tile-level wealth estimates with high resolution estimates of population density inferred from satellite imagery[41] to obtain population-weighted average wealth estimates for each canton, shown in Figure S3. For validation, we evaluate the canton-level correlation between average wealth from the satellite-based poverty map and average consumption in the 2018-2019 field survey (though note that the latter survey is not representative at the canton level). The Pearson correlation coefficient is 0.57, and the Spearman correlation coefficient is 0.52.

c. Mobile Phone Metadata

We obtain mobile phone metadata (call detail records, or CDR) from Togo's two mobile network operators for certain time periods in 2018-2021. We focus on three slices of mobile network data: October - December 2018, April - June 2019, and March - September 2020. The three-month periods in 2018 and 2019 are matched to households interviewed in the first and second wave of the field survey, respectively. The seven-month period in 2020 is matched to outcomes for individuals interviewed in the phone survey in September 2020.

Our CDR data contain the following information:

- *Calls:* Caller phone number, recipient phone number, date and time of call, duration of call, ID of the cell tower through which the call is placed.
- *SMS messages:* Sender phone number, recipient phone number, date and time of the message, ID of the antenna through which the message is sent.
- *Mobile data usage:* Phone number, date and time of transaction, amount of data consumed (upload and download combined).
- *Mobile money transactions:* Sender phone number, recipient phone number (if peer-to-peer), date and time of the transaction, amount of transaction, and broad category of transaction type (cash in, cash out, peer-to-peer, or bill pay).

October-December 2018 and April-June 2019 CDR: Between October 1 and December 30, 2018, there were a total of 4.84 million unique mobile network subscribers between the two mobile phone networks (where a subscriber is any phone number that places at least one call or SMS network on a network). Between April 1 and June 31, 2019, there were a total of 4.89 million mobile network subscribers. We identify spammers on the network as any phone number that placed an average of over 100 calls or 100 SMS messages per day, and remove any transactions associated with these numbers from our dataset. We remove 232 spammers in the 2018 time period and 162 spammers in the 2019 time period. In the 2018-2019 CDR, we observe only calls, SMS messages, and mobile money transactions (we do not observe mobile data usage).

March-September 2020 CDR: For data between March 1 and September 30, 2020, we observe a total of 5.83 million mobile network subscribers (note that this subscriber population does not necessarily reflect a 19% increase in subscribers from 2018-2019, since the slice is seven months rather than three months and there is significant month-to-month churn in subscribers). We identify spammers as described above, resulting in the removal of transactions associated with 107 spammers from the 2020 CDR dataset. In the 2020 CDR, we observe calls, SMS messages, mobile data usage, and mobile money transactions.

Featurization: For each subscriber observed on the network in each of the three time periods, we calculate a set of 857-1,042 "CDR features" that describe aspects of the subscriber's mobile phone behavior. These include:

- *Call and SMS features:* We use open-source library bandicoot[42] to produce around 700 features relating to the calls and SMS messages each subscriber places and receives. These range from general statistics (e.g. number of calls/SMS messages, balance of incoming vs. outgoing transactions), to social network characteristics (e.g. number and diversity of contacts), to measures of mobility based on cell tower locations (e.g. number of unique towers, radius of gyration).
- *Location features:* Based on the locations of each of the cell towers in Togo, we calculate information about where each subscriber places their transactions. Specifically, we calculate the number and percentage of calls placed in each of Togo's 40 prefectures, and the number of unique antennas, cantons, prefectures, and regions that each subscriber visits.
- *International transaction features:* Using country codes associated with phone numbers, we calculate the number of outgoing international transactions, separately for calls and SMS messages. We also calculate the total time spent on outgoing international calls.
- *Mobile money features:* For each of four variables relating to transaction size --- transaction amount, percent of balance, balance before transaction, and balance after transaction --- we calculate the mean, median, minimum, and maximum, separately for incoming and outgoing mobile money transactions. We also calculate the total transaction count for each subscriber (separately for incoming and outgoing) and the total number of unique mobile money contacts (separately for incoming and outgoing). We perform these calculations for all transactions together, as well as separately by transaction type (cash in, cash out, peer-to-peer, bill payments, and other transactions).
- *Mobile data features:* We calculate the total, mean, median, minimum, and maximum mobile data transaction for each subscriber, as well as the standard deviation in transaction size. We also calculate the total number of mobile data transactions and the number of unique days on which data is consumed. Note that mobile data features are only calculated for the 2020 CDR period, as our 2018-2019 CDR does not include mobile data records.

- *Operator:* In our feature dataset we include a dummy variable for which of the two mobile network operators each subscriber is associated with.

Matching survey and CDR datasets: Using phone number collected in surveys, we match survey observations to CDR features. As noted in Methods §3, there are 4,618 households in the 2018-2019 field survey that provide a phone number, of which 4,171 match to CDR (90% of households with phone numbers, and 68% of households overall). We match households surveyed in the first survey wave to features generated in the October-December 2018 CDR period, and households surveyed in the second survey wave to features generated in the April-June 2019 CDR period. Since the 2020 survey was sampled based on the CDR dataset, all 8,915 observations in the 2020 survey dataset are matched to CDR.

d. Data Privacy Concerns

The CDR data we obtained for each subscriber contain personally identifying information (PII) in the form of the subscriber's phone number (it does not contain the individual's name, address, or other PII), as well as other potentially sensitive information such as data about the subscriber's network and cell tower locations. To protect the confidentiality of these data, we pseudonymized the CDR prior to analysis by hash-encoding each phone number into a unique ID. The data are stored on secure university servers to which access is limited based on a data management plan approved by U.C. Berkeley's Committee for the Protection of Human Subjects.

We obtained informed consent from all research subjects in the phone survey prior to matching CDR records to survey responses. However, there are still open concerns around the use of CDR by bad actors, particularly as even pseudonymized datasets can frequently be de-anonymized for a subset of observations.[43,44] Active research on applying the guarantees of *differential privacy* to CDR datasets and associated machine learning models holds promise for balancing the utility of CDR data with privacy concerns.[45,46] For additional discussion of these considerations, see Methods §6.

4. Predicting poverty from mobile phone data

a. Machine Learning Methods

We follow the machine learning methods described in prior work[13,14,20] to train models that predict poverty from CDR features. Specifically, we train a gradient boosting regressor with Microsoft's LightGBM for the two matched survey-CDR datasets separately. We tune hyperparameters for the model over 3-fold cross validation, with parameters chosen from the following grid:

- *Winsorization of features:* {No winsorization, 1% limit}
- *Minimum data in leaf:* {10, 20, 50}
- *Number of leaves:* {5, 10, 20}
- *Number of estimators:* {20, 50, 100}
- *Learning rate:* {0.05, 0.075, 0.1}

We train and evaluate the model over 5-fold cross validation, with hyperparameters tuned independently on each fold, to obtain out-of-sample estimates of accuracy and out-of-sample predictions of poverty for each observation in our matched survey datasets. We then re-train the model on all survey data (for each of the two datasets separately), record feature importances (the total number of times a feature is split on over the entire forest), and use the final model to generate wealth predictions for every subscriber on the mobile phone network during the relevant time period.

We experiment with training models in this way for each of the relevant poverty outcomes: consumption expenditure, PMT, and asset index for the 2018-2019 field survey dataset and PMT and asset index for the 2020 phone survey dataset. Evaluations of model accuracy are found in Table S4 and feature importances for each model are shown in Figure S5. The correlation between the phone-based poverty predictions and a traditional PMT is 0.41, as trained and evaluated on the 2020 phone survey dataset (Table S4, Panel C). When trained and evaluated using the national 2018-2019 household survey with consumption data, the correlation between the phone-based poverty predictions and consumption is 0.46 (Table S4, Panel A).

We note that in examining the feature importances, location-related features (number and percent of calls placed in each prefecture of the country) are very important. The correlation between phone-based poverty predictions using only these location features and a standard PMT is 0.35 when trained and evaluated with the 2020 phone survey. The correlation between the location-only phone-based poverty predictions and consumption is 0.42 when trained and evaluated with the 2018-2019 field survey, only a small drop in accuracy from the full phone-based poverty scores. Other important features in the full phone-based poverty scores relate to nighttime calling behavior, mobile data usage, and mobile money usage (see Figure S5).

b. Parsimonious Method: "Phone Expenditure"

In addition to the machine learning method for wealth prediction described above, we are interested in the performance of an intuitive, parsimonious method for approximating poverty with CDR. We focus on a measure of "phone expenditure" on the basis of costs of all calls placed and SMS messages sent by each subscriber. We apply standard rates for calls and SMS messages in Togo: 30 CFA (USD 0.06) to send an SMS message and 50 CFA (USD 0.09) per minute of call time.[xiii] We use these prices to back out the (approximate) amount spent by each subscriber from their outgoing mobile phone transaction logs. We find that the "phone expenditures" method is significantly less accurate than the ML-based method, with a correlation of 0.13 with both the 2020 phone survey PMT and the 2018-2019 household survey's consumption measure (Table S4, Panels A and C).

[xiii] These prices represent a typical Togolese phone plan, though there is considerable diversity in special promotions and friends-and-family plans available from Togo's two mobile phone operators, Moov and Togocom.

5. Targeting evaluations

a. Experimental Design

We simulate CDR-based and counterfactual targeting methods for reaching the poorest individuals in Togo, using the two survey datasets described in Methods §3.a. For the 2018-2019 in-person survey dataset, the ground-truth wealth measure is consumption expenditure: we evaluate how well proxy measures of poverty reach those with the lowest consumption. For the 2020 phone survey dataset, the ground-truth wealth measure is the proxy-means test (since consumption information, which is more complicated and time-consuming to enumerate, was not collected in the phone survey).

Our evaluations measure how effectively several different targeting methods, described below, are at reaching the poorest individual mobile phone owners in each of the two survey populations. We focus on *individuals* rather than *households* because the Novissi program was designed and paid as an individual benefit. While social assistance programs in other countries sometimes determine benefits based on household composition, there is no notion of a household unit in the Novissi program (in part because the government does not possess data that links individuals to households)[xiv].

Likewise, our focus on mobile phone owners reflects the fact that the Novissi system in Togo distributed payments via mobile money; as such, anyone without access to a phone could not receive benefits irrespective of the targeting method – see Methods §f for a discussion of exclusion errors resulting from this constraint. In practice, this only affects the analysis using the 2018-2019 in-person survey, where 4,171 of 6,171 respondents provided a valid phone number. For analysis using the 2020 phone survey, we include all respondents, since every respondent had access to a phone. Future work could compare phone-based targeting to counterfactual targeting methods that could be implemented in-person, and thus account for exclusion errors resulting from phone ownership.

b. Targeting Methods and Counterfactuals

Our evaluations use the two survey datasets to measure the targeting performance of three targeting methods that were feasible when implementing the Novissi program: geographic blanketing (targeting everyone in certain geographies), occupation-based targeting (targeting everyone in certain occupation categories), and phone-based targeting.

With *geographic targeting*, the primary counterfactual approach considered by the government of Togo in implementing its rural assistance program, we assume that the program would target

[xiv] Note that in the 2020 phone survey the unit of observation is the individual, while in the 2018-2019 field survey the unit of observation is the household: in practice, this means that our simulations with the 2018-2019 field survey dataset reflect a program that would provide benefits only to heads of household, and we do not account for household size in considering exclusion errors or social welfare. Future work could model phone-based targeting on a household basis by collecting phone numbers for all household members and calculating aggregate benefits assigned to each household; given survey data limitations this analysis is beyond the scope of this paper.

geographic units in order from poorest to wealthiest, and that all individuals in targeted units would be eligible for benefits. We report results from geographic targeting at the prefecture level (admin-2 region, using survey-based estimates of wealth) and canton level (admin-3 region, using satellite-based estimates of wealth). See Figure S3 and Methods §3.b for the poverty maps used for geographic targeting.

In *occupation-based targeting*, we first evaluate the effectiveness of targeting *informal workers*, which is the eligibility criteria used by Novissi when it was first launched in April 2020, and which served as the basis for paying roughly 500,000 urban residents. In practice, this involves categorizing the occupation of every individual respondent in both surveys as either formal or informal (including unemployed), applying the same definition of informality that was used by the Novissi program. In the simulations, informal workers are targeted first (in random order if there are more informal workers than can receive benefits) and formal workers are targeted last (also in random order, if the available benefits exceed the number of informal workers).

We also develop and test a hypothetical occupation-based approach, which we refer to as *optimal occupation-based targeting*, which assumes that the policymaker had high-quality consumption data on the consumption of workers in each occupation and used that information to target the poorest occupations first. While this approach was not considered in Togo's pandemic response, it was feasible with the data sources available in Togo at the time, and represents an upper-bound on the performance of a hypothetical occupation-based targeting system. We simulate this optimal occupation-based approach by calculating the average consumption of each occupation in the 2018-2019 field survey; occupations are then targeted in order of increasing average consumption. The average consumption of each occupation category is shown in Table S3. Note that since agricultural workers are the poorest category and make up 29% of the observations in the 2018-2019 field survey dataset and 41% of the observations in the 2020 phone survey dataset, in practice the precision and recall metrics reported in our targeting simulations reflect systems of occupation-based targeting that would prioritize agricultural workers only.

Of primary interest in the targeting evaluation is the performance of the targeting approaches based on *mobile phone data*. The *phone-based (ML)* approach is the one described in the main text, which uses machine learning to construct a poverty score from rich data on mobile phone use and prioritizes the individuals with the lowest poverty scores. For reference, we also calculate the performance of a more parsimonious *phone (expenditures)* model, which prioritizes the individuals with the smallest total phone expenditures (Methods §4.b).

For completeness, our simulations also include results from targeting methods that were not feasible for the Novissi program, as the data required to implement those methods were not available when Novissi was launched (though Togo plans to create a foundational unique ID system and comprehensive social registry in 2022). In particular, we simulate targeting using an *asset-based wealth index*, constructed as described in Methods §3.a. For the hypothetical national simulations using the 2018-2019 field survey dataset, we also simulate targeting using a *poverty probability index* (PPI) and *proxy-means test (PMT)*. We cannot simulate PPI or PMT-based targeting using the 2020 phone survey since the necessary data were not collected.

An important caveat is that the PMT that we use in the 2018-2019 survey is "perfectly calibrated" in the sense that it is both trained and evaluated on the same sample. In real-world settings, the predictive accuracy of a PMT declines as the time increases between the time of calibration and the time of application[22,23]. As such, the performance of the PMT we report is likely an upper-bound of the performance of a real-world PMT.

For the PMT in the 2018-2019 field survey dataset, as well as for CDR-based wealth estimates in both datasets, predictions are produced out-of-sample over cross validation so that they can be fairly evaluated in targeting simulations. Specifically, in each case, the training dataset is divided into 10 cross validation folds; the machine learning model is trained on 9 of the 10 folds and used to produce predictions for the final fold. The training-and-prediction regime is repeated for all 10 folds.

c. Measures of Targeting Quality

For each targeting method, we calculate two "threshold-agnostic" metrics of targeting accuracy – metrics that capture relationships between continuous measures of poverty rather than focusing on accuracy for targeting a specific portion of the population. These are:

- *Spearman correlation coefficient:* Spearman's rank correlation coefficient is the Pearson correlation between the rank values of the true and proxy measures of poverty. We focus on the Spearman correlation rather than standard Pearson correlation as a measure of targeting quality because targeting concerns itself only with the ordering of observations according to poverty. Spearman's correlation coefficient is calculated as follows:

$$\rho = \frac{6\sum_{i=1}^{N}(r_i - \hat{r}_i)^2}{N(N^2 - 1)}$$

  where N is the total number of observations, r_i is the rank of observation i according to the ground truth poverty measure, and $\hat{r}_i$ is the rank of observation i according to the proxy poverty measure.

- *ROC curves and Area Under the Curve (AUC):* Following Hanna & Olken (2018)[3], we trace Receiver Operator Characteristic (ROC) curves that describe the quality of a targeting method at counterfactual targeting thresholds (Figure S6, left figures). At each counterfactual targeting threshold T we simulate targeting T% of observations according to the proxy poverty measure in question and calculate the true positive rate (TPR) and false positive rate (FPR) of the classifier with respect to reaching the T% poorest according to the ground-truth poverty measure. By varying T from 0% to 100%, we construct the ROC curves shown in Figure S6. The area under the curve (AUC) is used to summarize the targeting quality, with a random targeting method achieving an AUC of 0.5 and perfect targeting an AUC of 1. For convenience, we also include "Coverage vs. Recall" figures (right figures of Figure S6) that show how program recall varies as the eligible percentage of the population increases. Note that since recall is another name for the true positive rate, panels b and d represent a rescaling of the ROC curves in panels a and c.

We then focus on analyzing aid programs that target the poorest 29% of individuals as the Novissi aid program aimed to reach the poorest 29% of registrants in eligible cantons. The 29th percentile corresponds to a consumption threshold of USD 1.17/day in the 2018-2019 field survey dataset, and a PMT threshold of USD 1.18/day in the 2020 phone survey dataset. We calculate the following metrics to describe how accurately targeting the poorest 29% according to each targeting method reaches (1) the 29% truly poorest, (2) those below the international poverty line of USD 1.90/day (57% of observations in the 2018-2019 field survey, and 76% of observations in the 2020 phone survey), and (3) those below the extreme poverty line, which was defined as three-quarters of the poverty line, or USD 1.43/day (41% of observations in the 2018-2019 field survey, and 53% of observations in the 2020 phone survey):

- *Accuracy:* Classification accuracy measures the proportion of observations that are identified correctly (targeted observations that are poor according to the ground-truth poverty measure, and non-targeted observations that are not poor according to the ground-truth wealth measure). $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$
- *Precision:* Precision measures the proportion of targeted observations that are poor according to the ground-truth poverty measure. $Precision = \frac{TP}{TP+FP}$
- *Recall:* Recall measures the proportion of all poor observations that are reached by a given targeting method. $Recall = \frac{TP}{TP+FN}$

Note that the poverty lines are applied to consumption expenditure in the 2018-2019 field survey dataset, and to the proxy-means test estimates in the 2020 phone survey dataset.

d. Social Welfare

Using the two matched survey-CDR datasets, we calculate aggregate utility under each of the targeting methods using a social welfare function. Following Hanna & Olken (2018)[3] we rely on constant relative risk-aversion (CRRA) utility, which models individual utility as a function of pre-transfer consumption and transfer size:

$$U = \frac{\sum_{i=0}^{N} (y_i + b_i)^{1-\rho}}{1-\rho}$$

Where N is the population size, y_i is the consumption of individual i, and b_i the benefits assigned to the individual. Following Hanna & Olken (2018)[3]**,** we use a coefficient of relative risk-aversion $\rho = 3$. To reflect the policy design of the Novissi program, we assume that all beneficiaries who receive a benefit receive the same value $b_i = b$.[xv] To construct the social welfare curves, we:

- Calculate a total budget available for each of the two datasets. We focus on programs that have a budget size analogous to that of rural Novissi, which aimed to distributed approximately USD 4 million among the 154,238 program registrants, or USD 25.93 per

[xv] In principle, the benefit b_i paid to i could depend on characteristics of i, such as i's level of poverty. While such an approach would substantially increase total welfare, in practice it is much more difficult to implement.

voter. We therefore assign each dataset a total budget of USD 25.93N, where N is the total size of the dataset.

- Simulate targeting T% of observations on the basis of each of our counterfactual targeting approaches.
- Assign equal benefits to each of the targeted observations, with the budget divided evenly among targeted observations (so lower targeting thresholds T correspond to more benefits for targeted individuals).
- Calculate aggregate utility by summing over benefits and consumption for each individual with the CRRA utility function. Note that non-targeted individuals are included in the welfare calculation; they are merely assigned 0 benefits. For the 2018-2019 field survey dataset we use consumption expenditure for y_i for the 2020 phone survey dataset we use the PMT estimates.
- By varying T between 0% and 100% of observations targeted, we trace out the social welfare curves shown in Figure 3.

e. Fairness

We are interested in auditing our targeting methods for fairness across sensitive subgroups. Note that notions of parity and fairness are debated in machine learning and policy communities: Kleinberg et al. (2016)[27] describe how the three most popular parity criteria --- demographic parity (benefits assigned to subgroups proportionally to their size), threshold parity (use of the same classification threshold for all subgroups), and error rate parity (equal classification error across subgroups) --- are in tension with one another. Moreover, Noriega et al. (2020)[26] describe how tensions over parity criteria, prioritized subgroups, and positive discrimination lead to complicated prioritization compromises in the administration of targeted social protection programs.

Here we focus on two targeting-specific parity criteria:

- *Demographic parity:* A targeting method satisfying demographic parity will assign benefits to a subgroup proportionally to the subgroup's presence in the population of interest. We evaluate demographic parity among the poor: that is, we compare the proportion of each subgroup living in poverty (below the 29th percentile in terms of consumption) to the proportion of each subgroup that is targeted (below the 29th percentile in terms of the proxy poverty measure used for targeting).

$$DP = \frac{True\ Positives + False\ Positives}{N} - \frac{True\ Positives + False\ Negatives}{N}$$

- *Normalized rank residual:* We are interested in whether certain subgroups are consistently ranked higher or consistently ranked lower than they "should" be by the counterfactual targeting approaches. We therefore compare the distributions of rank residuals across subgroups and targeting methods:

$$RR_i = \frac{\hat{r}_i - r_i}{N}$$

where $\hat{r}_i$ is the poverty rank of individual *i* according to the proxy poverty measure and r_i is the poverty rank of individual *i* according to the ground-truth poverty measure.

We focus on seven dimensions for parity: gender, ethnicity, religion, age group, disability status, number of children, and marital status. We also evaluate parity across whether an individual is "vulnerable," where vulnerability is defined as one of the following traits: {female, over age 60, has a disability, has more than five children, is single}. We conduct this analysis with only the 2018-2019 field survey dataset, as these demographic variables were not all collected in the 2020 phone survey. We use demographic information from the head of household.

f. Program Exclusions

We present information on sources of exclusion from the Novissi program that are not inherently related to targeting. These statistics are drawn from diverse sources of administrative and survey data, specifically:

- *Voter ID penetration:* According to government administrative datasets, 3,633,898 individuals were registered to vote in Togo by late 2019. The electoral commission of Togo reports that this corresponds to 86.6% of eligible adults. While the total adult population in Togo is hard to pin down (the last census was in 2011), Togo's national statistical agency (https://inseed.tg/) estimates that there are 3,715,318 adults in Togo, whereas the United Nations estimates 4.4 million adults in Togo[47], implying a voter ID penetration rates of 82.6% or 97.8%.
- *Phone penetration:* In the 2018-2019 field survey, 65% of individuals report owning a mobile phone (Figure S9a); 54% of individuals in rural areas report owning a mobile phone. 53% of women and 79% of men report owning a mobile phone; 85% of households nationally include at least one individual who owns a phone (Figure S9b). These household estimates likely represent a lower bound, given the steady increase in phone penetration between 2018 and 2020. The Togolese government estimates 82% SIM card penetration in the country (though some people may have multiple SIM cards)[xvi]. Based on data from the mobile phone companies, we observe 5.83 million unique active SIMs in Togo between March and September 2020.
- *Program awareness:* Since individuals had to register for the Novissi program to receive benefits, program advertising and population awareness was a key goal. The program was advertised via radio, SMS, field teams, and direct communication with community leaders at the prefecture and canton level. In total, 245,454 subscribers attempt to register for the program. Although we do not observe the prefecture and canton of subscribers who attempt but do not succeed in registering in our administrative data, we know that 87% of successful registrants are in cantons eligible for benefits. Assuming the rate is approximately the same for attempters, we expect that around 213,545 of the attempters

[xvi] "Programme digital de transferts monétaires en réponse à la COVID-19", accessed July 14, 2021 from http://www.fondation-farm.org/zoe/doc/colocnovissi.pdf

are in eligible cantons. The total voting population in eligible cantons is 528,562, for an estimated attempted registration rate of 40.40%.

- *Registration challenges:* Registration for the Novissi program required the completion of a short (5 question) USSD survey. Of the 245,454 subscribers that attempted to register for the program, 176,517 succeed, for a 71.91% rate of registration success.

Overlaps among sources of exclusion: The above sources of exclusion are not independent and are therefore not cumulative. For instance, individuals who are not registered to vote may also be systematically less likely to have a mobile phone. For this reason, we use the 2020 phone survey dataset – restricted to respondents who report living in an eligible canton – to calculate overlaps in sources of exclusion to the poor, including Voter ID possession, program awareness, registration challenges, and targeting errors using the CDR-based targeting method. We cannot account for mobile phone ownership in this analysis since the 2020 survey was conducted over the phone. The resulting attrition table is available in Table S8.

6. Limitations and Concerns

While mobile phone data can create new options for the accurate targeting of humanitarian aid, there are several important limitations. A full discussion of the social, political, and ethical implications of these issues has been the focus of prior work and is beyond the scope of this article[48–52]; we nonetheless highlight a few key issues that we believe require careful consideration before these methods can be implemented in a policy environment:

Phone ownership and access: As discussed in Methods §5.f, many individuals in LMICs do not own mobile phones. Thus, any targeting system based on mobile phone data may exclude those without phones from receiving program benefits. In the case of the Novissi program, the government used the mobile money system to disburse the cash transfers as a way to minimize human contact during the pandemic. Thus, in Togo, the use of phone data for targeting did not create further exclusions than those already created by the payment mechanism. However, in general, incomplete mobile phone access highlights the need to allow for alternative pathways for individuals to register and receive benefits, and to create additional mechanisms for appeals, grievance redress mechanisms, and manual enrollment.

Data privacy: Mobile phone metadata, even when pseudonymized, contains sensitive information. Methods §3.d describes several steps taken to protect the confidentiality of the data used in this project. More generally, special considerations arise when using personal data from vulnerable populations[53], and human rights doctrine emphasizes that any form of communications surveillance should be "necessary and proportionate"[54].

In implementing the approach described in this paper, we developed an IRB protocol, as well as a data management plan, that was approved by U.C. Berkeley's Committee for the Protection of Human Subjects. We followed principles of data minimization to limit the data collected and stored, and implemented organizational safeguards to restrict access to data. As an example, only IRB-approved researchers ever received access to CDR; data from the phone companies were shared with neither the Government of Togo nor GiveDirectly. Even the poverty scores derived from the phone data were restricted to IRB-approved researchers; the only data the government received was the list of SIM cards belonging to eligible beneficiaries below the targeted poverty threshold.

Future projects using mobile phone data for targeting should ensure that principles of data minimization and data sunsetting restrict the use of sensitive data to social protection objectives and limit the potential for "function creep."[55] Further research on applying the guarantees of differential privacy to mobile phone metadata[45,46] or implementing federated learning systems[56] could reduce the risk of data misuse or central data breaches.

Data access and consent: The fact that our approach requires access to mobile phone data owned by private companies poses an obstacle to the immediate and widespread use of such data for targeting humanitarian aid. There now exist several general frameworks and recommendations to facilitate the use of CDR in humanitarian applications[49,57]. Yet such frameworks are still nascent, and without careful consideration may exclude important stakeholders and perspectives[52]; they also widen the scope for private companies to influence humanitarian and development decisions[58]. There also exist many ethical frameworks rely on informed consent from participants for the use of personal data, including digital data such as CDR.[59,60] Future programs should consider how consent pathways can be integrated with phone-based targeting, including opt-in (calculating poverty scores only after consent is provided) and opt-out (scrubbing data if consent is not provided at the time of registration) options.

Data Representativity: To train the machine learning models, ground truth measures of consumption and wealth were collected using in-person and phone surveys. Since response rates were imperfect in the phone survey, we reweighted survey observations to make the training data more representative of all mobile subscribers (Methods §3.a). However, there are limits to the representativity of our training data, as dynamics of phone ownership and phone sharing vary across population subgroups (Figure S9), and reweighting is an imperfect proxy (Methods §3.a).

To test for systematic bias based on data representativity, we perform ex-post audits to limit the likelihood that the trained models systematically disadvantage specific subgroups of the population (Methods §5.e), and find that the phone-based targeting method is no more biased that counterfactual targeting approaches. We believe such audits are essential to future work on wealth prediction and targeting based on nontraditional data. Audits could be improved with additional context-specific research about which sub-populations are at the greatest risk for systematic exclusion (for example, in this paper we test for bias across age groups, genders, ethnicities, and more), and on considering alternative definitions for bias and fairness.[26,27]

Manipulation and gaming: When mobile phone data are used to determine eligibility for social benefits, individuals have incentives to strategically alter their behavior in order to "game" the system. This dilemma is not unique to phone-based targeting; it is a key consideration in the design of any targeting mechanism[61,62], and one that affects traditional proxy means tests and poverty scorecards[63,64]. However, recent evidence suggests that such distortionary effects may be limited[65], and complex eligibility criteria (such as the gradient boosting procedure described in Methods §4) should limit the scope for such gaming[66]. With Novissi in Togo, the one-off nature of the program likely eliminated most scope for strategic behavior; however, if such an approach were used continuously over time, alternative “manipulation-proof” approaches to machine learning may be more appropriate[67].

Acknowledgements

Isabel Onate Falomir, Shikhar Mehra, Suraj Nair, Adrian Dar Serapio, Nathaniel Ver Steeg, and Rachel Warren provided invaluable research assistance on this project. This project would not have been possible without the dedication of our project partners in Togo, especially Minister Cina Lawson, Shegun Bakari, Stanislas Telou, Leslie Mills, Kafui Ekouhoho, Morlé Koudeka, and Attia Byll. The team at GiveDirectly was instrumental in implementing the Novissi expansion studied in this paper (especially Han Sheng Chia, Michael Cooke, Kristen Lee, Alex Nawar, and Daniel Quinn). We thank Esther Duflo, Luis Encinas, Tina George, Rema Hanna, Ethan Ligon, and Ben Olken for helpful feedback. We are grateful for financial support from Google.org, data.org, the Center for Effective Global Action, the Jameel Poverty Action Lab, the Global Poverty Research Lab at Northwestern University, and the World Bank, which financed the phone surveys and data collection under the WURI program. Blumenstock is supported by NSF award IIS – 1942702. Authors retained full intellectual freedom in conducting this research, and as such all opinions and errors are our own.

Data Availability

The data used in this analysis include data that are available from public online repositories, data that are publicly available upon request of the data provider, and data that are not publicly available because of restrictions by the data provider.

Code Availability

The code used for these analyses will be made publicly available through the GitHub repository located at https://github.com/emilylaiken/togo-ml.

References

1. Egger, D. *et al.* Falling living standards during the COVID-19 crisis: Quantitative evidence from nine developing countries. *Science Advances* **7**, eabe0997 (2021).
2. Gentilini, U., Almenfi, M., Orton, I. & Dale, P. Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures. *World Bank Policy Brief* (2020).
3. Hanna, R. & Olken, B. A. Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries. *Journal of Economic Perspectives* **32**, 201–226 (2018).
4. Yeh, C. *et al.* Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications* **11**, 2583 (2020).
5. Bank, W. *World Development Report 2021: Data for Better Lives*. (The World Bank, 2021).
6. Jerven, M. *Poor numbers: how we are misled by African development statistics and what to do about it*. (Cornell University Press, 2013).
7. Lakner, C., Yonzan, N., Mahler, D., Aguilar, R. A. & Wu, H. *Updated estimates of the impact of COVID-19 on global poverty: Looking back at 2020 and the outlook for 2021*. https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty-looking-back-2020-and-outlook-2021.
8. Mirrlees, J. A. An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies* **38**, 175–208 (1971).
9. Serajuddin, U., Wieser, C., Uematsu, H., Dabalen, A. L. & Yoshida, N. *Data deprivation : another deprivation to end*. 1–24 http://documents.worldbank.org/curated/en/700611468172787967/Data-deprivation-another-deprivation-to-end (2015).
10. Coady, D., Grosh, M. & Hoddinott, J. Targeting outcomes redux. *The World Bank Research Observer* **19**, 61–85 (2004).
11. Jean, N. *et al.* Combining satellite imagery and machine learning to predict poverty. *Science* **353**, 790–794 (2016).
12. Chi, G., Fang, H., Chatterjee, S. & Blumenstock, J. E. Micro-Estimates of Wealth for all Low- and Middle-Income Countries. *In Submission* (2021).
13. Blumenstock, J. E., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**, 1073–1076 (2015).
14. Blumenstock, J. E. Estimating Economic Characteristics with Phone Data. *American Economic Review: Papers and Proceedings* **108**, 72–76 (2018).
15. Nichols, A. L. & Zeckhauser, R. J. Targeting Transfers through Restrictions on Recipients. *The American Economic Review* **72**, 372–377 (1982).
16. Grosh, M. E. & Baker, J. L. *Proxy means tests for targeting social programs*. (The World Bank, 1995). doi:10.1596/0-8213-3313-5.
17. Alatas, V., Banerjee, A., Hanna, R., Olken, B. A. & Tobias, J. Targeting the Poor: Evidence from a Field Experiment in Indonesia. *American Economic Review* **102**, 1206–40 (2012).
18. Blumenstock, J. Machine learning can help get COVID-19 aid to those who need it most. *Nature* **581**, (2020).
19. Lindert, K., Karippacheril, T. G., Caillava, I. R. & Chávez, K. N. *Sourcebook on the Foundations of Social Protection Delivery Systems*. (World Bank Publications, 2020).

20. Aiken, E., Bedoya, G., Coville, A. & Blumenstock, J. E. Targeting Development Aid with Machine Learning and Mobile Phone Data. *Working Paper* (2020).
21. McBride, L. & Nichols, A. Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. *World Bank Economic Review* **32**, 531–550 (2018).
22. Brown, C., Ravallion, M. & van de Walle, D. A poor means test? Econometric targeting in Africa. *Journal of Development Economics* **134**, 109–124 (2018).
23. Skoufias, E., Diamond, A., Vinha, K., Gill, M. & Dellepiane, M. R. Estimating poverty rates in subnational populations of interest: An assessment of the Simple Poverty Scorecard. *World Development* **129**, 104887 (2020).
24. Eubanks, V. *Automating inequality: How high-tech tools profile, police, and punish the poor*. (St. Martin's Press, 2018).
25. Barocas, S., Hardt, M. & Narayanan, A. *Fairness and Machine Learning*. (fairmlbook.org, 2018).
26. Noriega-Campero, A. *et al.* Algorithmic targeting of social policies: fairness, accuracy, and distributed governance. in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* 241–251 (ACM, 2020). doi:10.1145/3351095.3375784.
27. Kleinberg, J., Mullainathan, S. & Raghavan, M. Inherent Trade-Offs in the Fair Determination of Risk Scores. *arXiv:1609.05807 [cs, stat]* (2016).
28. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. Fairness Through Awareness. *arXiv:1104.3913 [cs]* (2011).
29. Alatas, V. *et al.* Self-Targeting: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy* **124**, 371–427 (2016).
30. Kleven, H. J. & Kopczuk, W. Transfer Program Complexity and the Take-Up of Social Benefits. *American Economic Journal: Economic Policy* **3**, 54–90 (2011).
31. Baker, J. L. & Grosh, M. E. Poverty reduction through geographic targeting: How well does it work? *World Development* **22**, 983–995 (1994).
32. Schady, N. R. Picking the Poor: Indicators for Geographic Targeting in Peru. *Review of Income and Wealth* **48**, 417–433 (2002).
33. Coady, D. P. The Welfare Returns to Finer Targeting: The Case of The Progresa Program in Mexico. *Int Tax Public Finan* **13**, 217–239 (2006).
34. Elbers, C., Fujii, T., Lanjouw, P., Özler, B. & Yin, W. Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics* **83**, 198–213 (2007).
35. Smythe, I. & Blumenstock, J. E. Geographic Micro-Targeting of Social Assistance with High-Resolution Poverty Maps. in *In Submission (KDD)* (2021).
36. Filmer, D. & Pritchett, L. H. Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application To Educational Enrollments In States Of India*. *Demography* **38**, 115–132 (2001).
37. Desiere, S., Vellema, W. & D'Haese, M. A validity assessment of the Progress out of Poverty Index (PPI)™. *Evaluation and Program Planning* **49**, 10–18 (2015).
38. Alderman, H. Do local officials know something we don't? Decentralization of targeted transfers in Albania. *Journal of Public Economics* **83**, 375–404 (2002).
39. Galasso, E. & Ravallion, M. Decentralized targeting of an antipoverty program. *Journal of Public Economics* **89**, 705–727 (2005).

40. Premand, P. & Schnitzer, P. Efficiency, Legitimacy, and Impacts of Targeting Methods: Evidence from an Experiment in Niger. *The World Bank Economic Review* (2020) doi:10.1093/wber/lhaa019.
41. Tiecke, T. G. *et al.* Mapping the world population one building at a time. *arXiv:1712.05839 [cs]* (2017).
42. De Montjoye, Y.-A., Rocher, L. & Pentland, A. S. bandicoot: a python toolbox for mobile phone metadata. *J. Mach. Learn. Res.* **17**, 6100–6104 (2016).
43. de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports* **3**, 1376 (2013).
44. Cecaj, A., Mamei, M. & Bicocchi, N. Re-identification of anonymized CDR datasets using social network data. in *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)* 237–242 (2014). doi:10.1109/PerComW.2014.6815210.
45. Alaggan, M., Gambs, S., Matwin, S. & Tuhin, M. Sanitization of Call Detail Records via Differentially-Private Bloom Filters. in *Data and Applications Security and Privacy XXIX* (ed. Samarati, P.) vol. 9149 223–230 (Springer International Publishing, 2015).
46. Mir, D., Isaacman, S., Caceres, R., Martonosi, M. & Wright, R. DP-WHERE: Differentially private modeling of human mobility. in 580–588 (2013). doi:10.1109/BigData.2013.6691626.
47. United Nations. 2019 Revision of World Population Prospects. *United Nations Department of Economic and Social Affairs* (2019).
48. Mann, L. Left to Other Peoples' Devices? A Political Economy Perspective on the Big Data Revolution in Development. *Development and Change* **49**, 3–36 (2018).
49. Kerry, C. F., Kendall, J. & de Montjoye, Y.-A. Enabling Humanitarian Use of Mobile Phone Data. *Brookings Issues in Technology Innovation* (2014).
50. Blumenstock, J. E. Don't forget people in the use of big data for development. *Nature* **561**, 170–172 (2018).
51. Ohlenberg, T. *AI in Social Protection – Exploring Opportunities and Mitigating Risks*. (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2020).
52. Abebe, R. *et al.* Narratives and Counternarratives on Data Sharing in Africa. in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 329–341 (Association for Computing Machinery, 2021). doi:10.1145/3442188.3445897.
53. Taylor, L. No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environ Plan D* **34**, 319–336 (2016).
54. Necessary and Proportionate. International Principles on the Application of Human Rights to Communications Surveillance. (2013).
55. Taylor, L. No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environ Plan D* **34**, 319–336 (2016).
56. Li, T., Sahu, A. K., Talwalkar, A. & Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine* **37**, 50–60 (2020).
57. Oliver, N. *et al.* Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* eabc0764 (2020) doi:10.1126/sciadv.abc0764.
58. Taylor, L. & Broeders, D. In the name of Development: Power, profit and the datafication of the global South. *Geoforum* **64**, 229–237 (2015).
59. Cate, F. H. & Mayer-Schönberger, V. Notice and consent in a world of Big Data. *International Data Privacy Law* **3**, 67–73 (2013).

60. Ioannidis, J. P. A. Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research. *The American Journal of Bioethics* **13**, 40–42 (2013).
61. Goodhart, C. *Monetary Relationships: A View from Threadneedle Street*. (University of Warwick, 1975).
62. Akerlof, G. A. The economics of 'tagging' as applied to the optimal income tax, welfare programs, and manpower planning. *The American economic review* **68**, 8–19 (1978).
63. Martinelli, C. & Parker, S. W. Deception and Misreporting in a Social Program. *Journal of the European Economic Association* **7**, 886–908 (2009).
64. Camacho, A. & Conover, E. Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy* **3**, 41–65 (2011).
65. Banerjee, A., Hanna, R., Olken, B. A. & Sumarto, S. *The (lack of) Distortionary Effects of Proxy-Means Tests: Results from a Nationwide Experiment in Indonesia*. http://www.nber.org/papers/w25362 (2018) doi:10.3386/w25362.
66. Finkelstein, A. E-Z Tax: Tax Salience and Tax Rates. *The Quarterly Journal of Economics* **124**, 969–1010 (2009).
67. Björkegren, D., Blumenstock, J. E. & Knight, S. Manipulation-Proof Machine Learning. *arXiv:2004.03865 [cs, econ]* (2020).

Figures

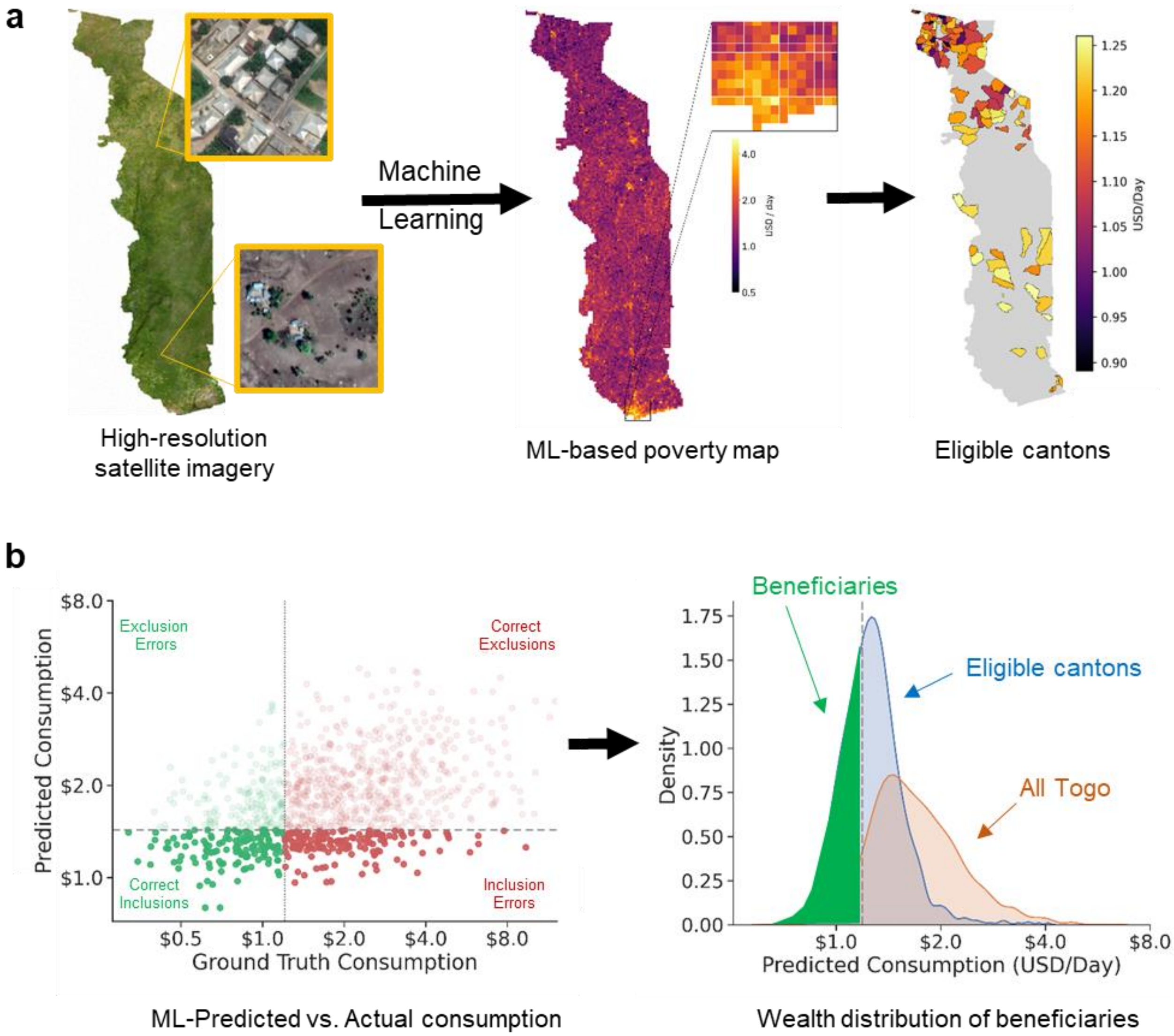

Figure 1 | Overview of targeting methodology for rural Novissi. a) *Regional targeting*. High-resolution satellite imagery (left) is analyzed with deep learning algorithms to generate poverty estimates for 1km^2 grid cells (middle), which are overlayed with population data to produce canton-level estimates of wealth. Individuals registered in the 100 poorest cantons (right) are eligible for benefits. **b)** *Individual targeting.* A machine learning algorithm is trained using representative survey data to predict consumption from features of phone use (Methods §4). The algorithm constructs poverty scores that are correlated with ground-truth measures of consumption (left). Subscribers who register for the program in targeted cantons with estimated consumption less than USD $1.25/day are eligible for benefits (right). The red distribution shows the predicted wealth distribution of the entire population of Togo; the blue distribution shows the predicted wealth distribution in the 100 poorest cantons; and the green section indicates the predicted wealth distribution of Novissi beneficiaries.

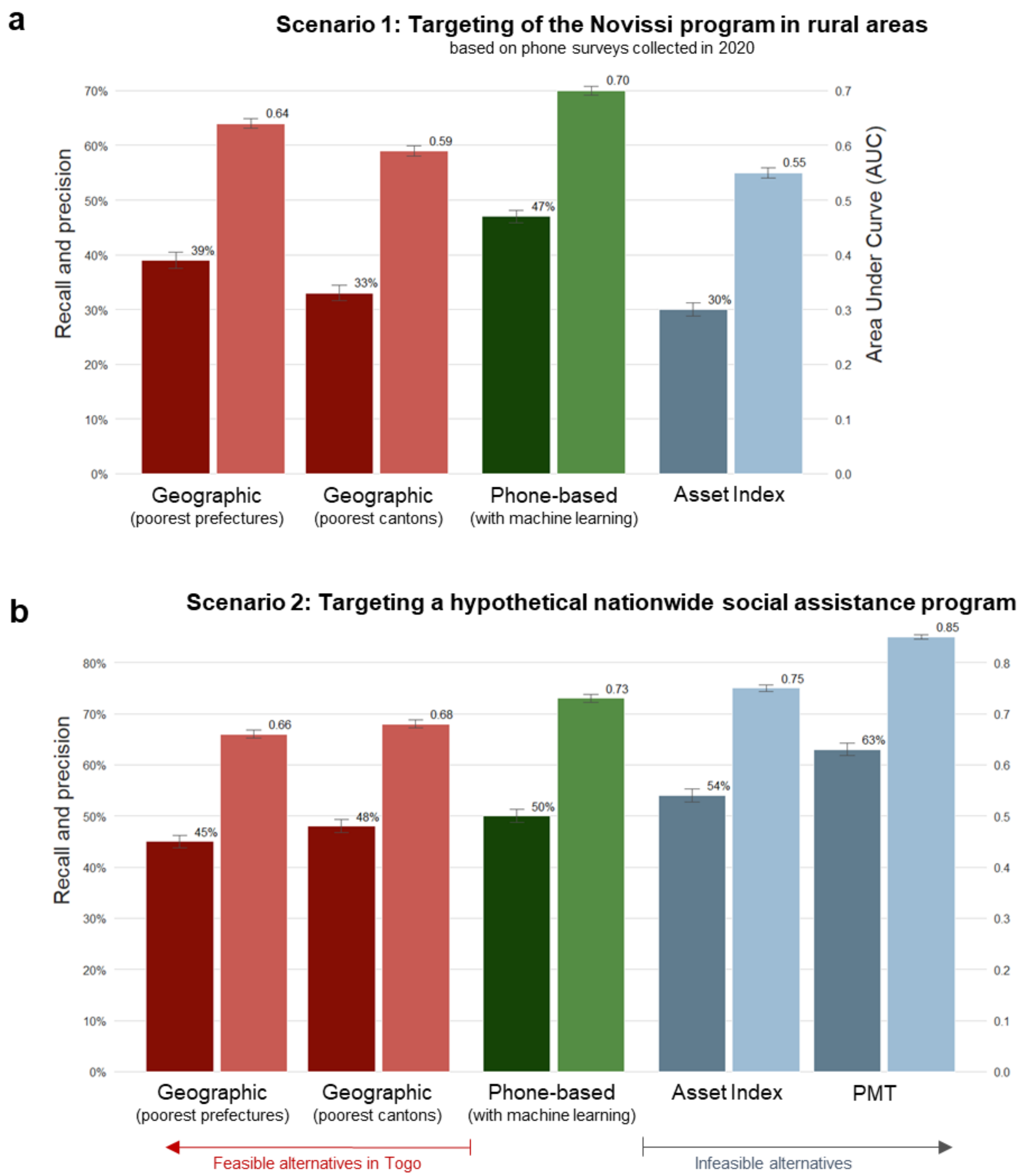

Figure 2 | Comparing Novissi targeting to alternatives. The performance of phone-based targeting (green) in comparison to alternative approaches that were feasible (red) and infeasible (grey) in Togo in 2020. Targeting is evaluated for **(a)** The actual rural Novissi program, which focused on Togo's 100 poorest cantons (using a 2020 survey representative of mobile subscribers in the 100 cantons, where PMT is a ground truth for poverty since consumption data was not collected in the phone survey); and **(b)** a hypothetical nationwide anti-poverty program (using a national field survey conducted in 2018-2019, where consumption is a ground truth for poverty). Darker bars indicate recall and precision (left axis); lighter bars indicate Area Under Curve (right axis). Standard deviations, produced from 1,000 bootstrap simulations, shown as whiskers. This figure highlights a subset of the results contained in Table 1.

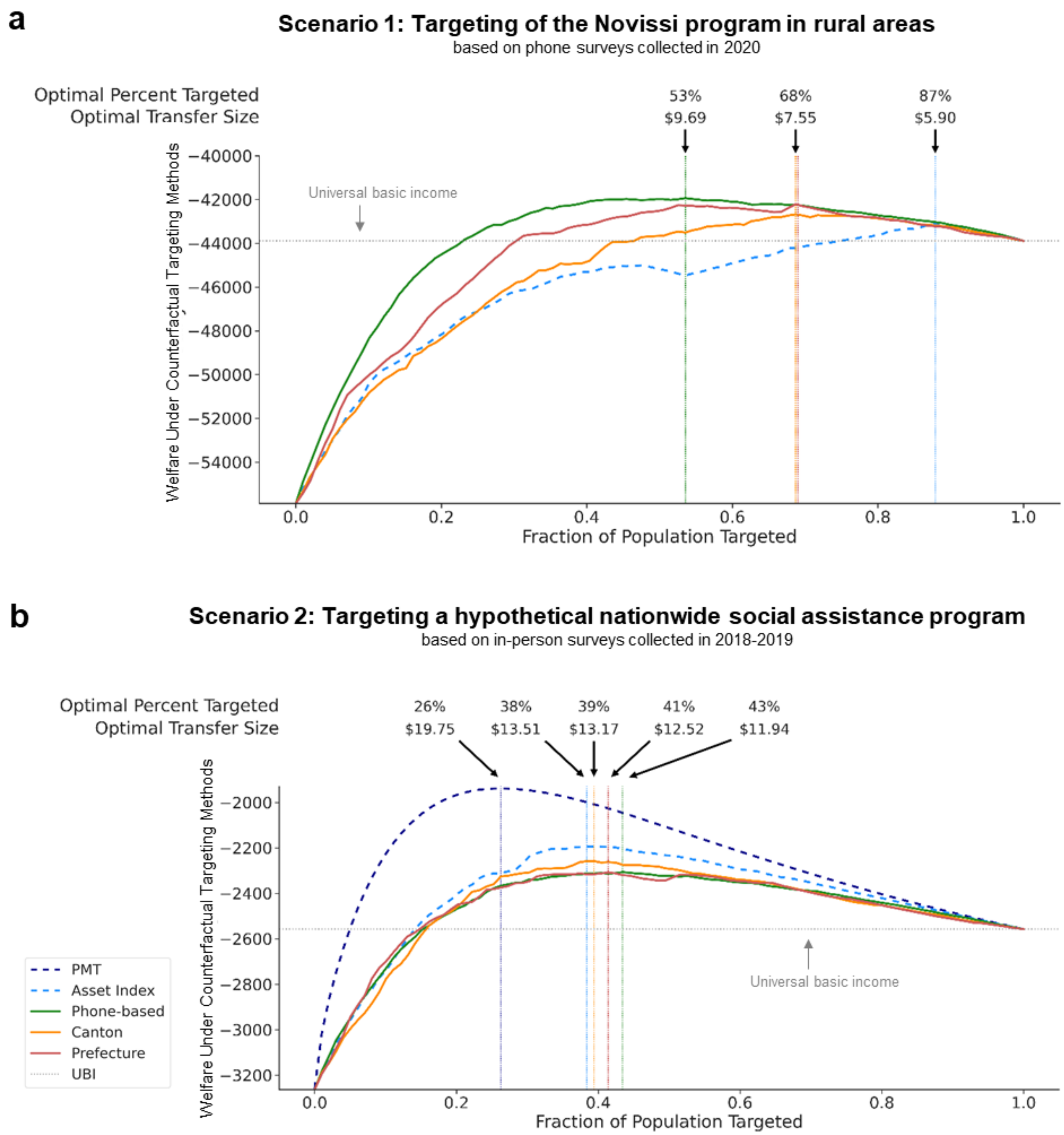

Figure 3 | Welfare analysis of different targeting mechanisms. Aggregate social welfare is calculated (assuming CRRA utility) under counterfactual targeting approaches. We assume a fixed budget of USD 4 million and a population of 154,238, with an equal transfer size for all beneficiaries. Utility curves for feasible targeting mechanisms are shown in solid lines; infeasible targeting mechanisms are shown in dashed lines. The horizontal dotted line indicates total social welfare for a universal basic income program that provides (very small) transfers to the entire population; vertical dotted lines indicate the targeting threshold and associated transfer size that maximizes social welfare for each targeting mechanism. Targeting is evaluated for **(a)** an anti-poverty program in Togo's 100 poorest cantons; and **(b)** a hypothetical nationwide anti-poverty program.

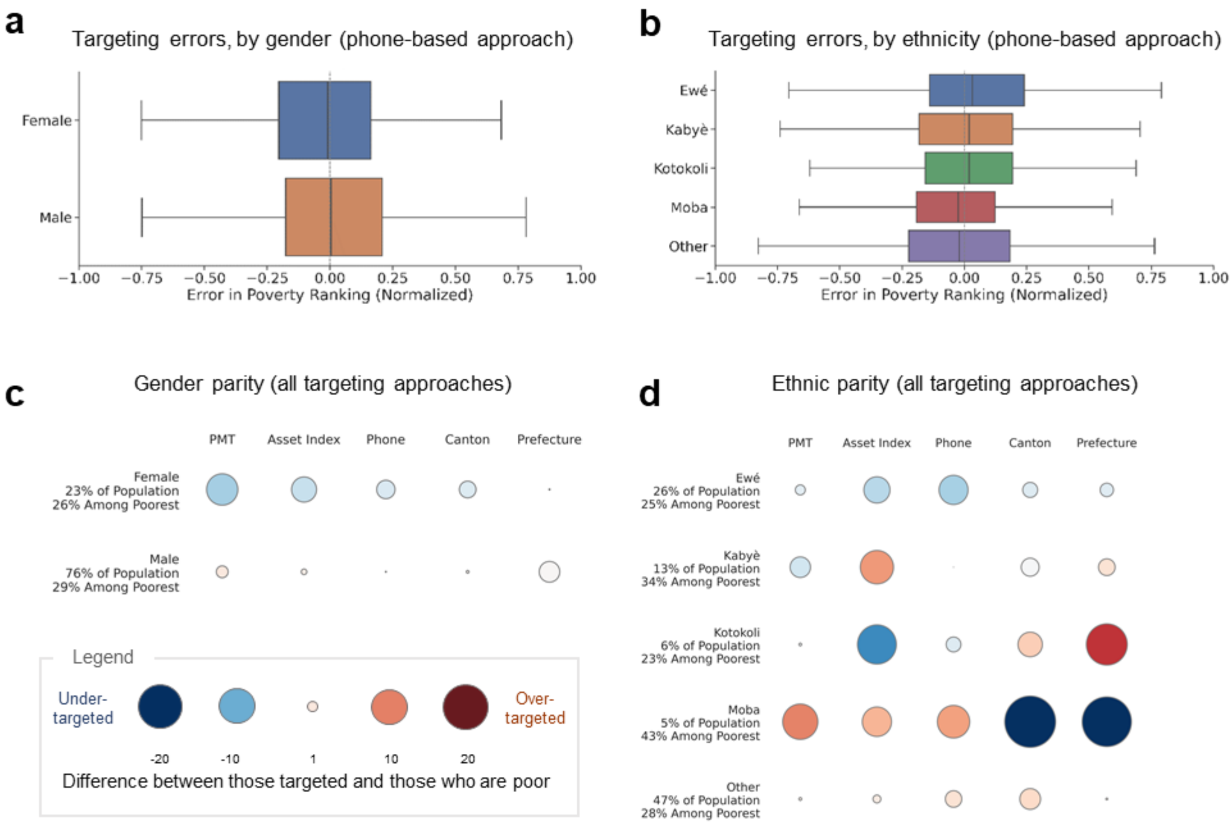

Figure 4 | Fairness of targeting for different demographic subgroups. Above: Distributions of differences between ranking according to predicted wealth from the ML approach and ranking according to true wealth (using the 2018-2019 field survey matched to CDR, $N = 4{,}171$), dissagregated by gender **(a)** and ethnicity **(b)**. Left-skewed bars indicate groups that are consistently under-ranked; right-skewed bars indicate groups that are consistently over-ranked. **Below:** Evaluation of demographic parity across subgroups by comparing the proportion of a subgroup targeted under counterfactual approaches to the proportion of the subgroup that falls into the poorest 29% of the population (using data from the 2018-2019 field survey matched to CDR, $N = 4{,}171$). Bubbles show the percentage point difference between the proportion of the subgroup that is targeted and the proportion that is poor according to ground-truth data. Large red bubbles indicate groups that are over-targeted; large blue bubbles indicate groups that are under-targeted.

	Targeting Novissi in rural Togo Based on 2020 Phone Survey (N = 8,915)				Hypothetical nationwide program Based on 2018-2019 Field Survey (N = 4,171)			
	Spearman	AUC	Accuracy	Precision & Recall	Spearman	AUC	Accuracy	Precision & Recall
Panel A: Targeting methods considered by the Government of Togo in 2020								
Prefecture (Admin-2 regions)	0.30 (0.017)	0.64 (0.008)	65% (0.87%)	39% (1.51%)	0.34 (0.017)	0.66 (0.008)	68% (0.74%)	45% (1.27%)
Canton (Admin-3 regions)	0.19 (0.019)	0.59 (0.009)	61% (0.78%)	33% (1.35%)	0.39 (0.016)	0.68 (0.008)	70% (0.71%)	48% (1.23%)
Phone (Expenditures)	0.13 (0.020)	0.57 (0.010)	60% (0.71%)	32% (1.23%)	0.26 (0.017)	0.63 (0.009)	65% (0.81%)	40% (1.40%)
Phone (Machine Learning)	0.38 (0.017)	0.70 (0.009)	69% (0.87%)	47% (1.18%)	0.45 (0.015)	0.73 (0.007)	71% (0.74%)	50% (1.28%)
Panel B: Common alternative targeting methods that could not be implemented in Togo in 2020								
Asset Index	0.10 (0.018)	0.55 (0.009)	60% (0.48%)	30% (0.83%)	0.51 (0.014)	0.75 (0.007)	74% (0.69%)	54% (1.19%)
PPI	*[data not available]*				0.63 (0.011)	0.81 (0.006)	77% (0.73%)	60% (1.25%)
PMT	*[data not available]*				0.72 (0.009)	0.85 (0.005)	78% (0.70%)	63% (1.20%)
Panel C: Additional counterfactual targeting methods that were feasible in Togo in 2020								
Random	0.00 (0.021)	0.50 (0.082)	59% (0.74%)	30% (0.26%)	0.00 (0.019)	0.50 (0.010)	59% (0.79%)	29% (1.36%)
Occupation (As implemented)	-0.11 (0.019)	0.45 (.007)	55% (0.62%)	22% (1.07%)	-0.09 (0.019)	0.46 (0.095)	56% (0.53%)	24% (0.91%)
Occupation (Optimally designed)	0.25 (0.016)	0.61 (0.008)	66% (0.58%)	41% (1.00%)	0.41 (0.016)	0.69 (0.008)	72% (0.72%)	52% (1.25%)

Table 1 | Performance of targeting mechanisms. Targeting performance using mobile phone data and machine learning (highlighted) in comparison to counterfactual targeting strategies. The "true poor" are those who, according to survey data, are in the poorest 29% of the population (the 29% threshold reflects the budget constraint of the rural Novissi expansion). The first four columns evaluate targeting with a 2020 phone survey representative of subscribers in Togo's 100 poorest cantons, using a PMT as ground truth for poverty since consumption data were not collected. The last four columns evaluate targeting using nationally representative household survey data collected in 2018-2019, using consumption as a ground truth. Panel A compares the phone-based PMT (highlighted) to alternative targeting methods that the Government of Togo considered prior to expanding Novissi to rural areas. Panel B shows the performance of targeting methods that are commonly implemented but were infeasible in Togo at the time. Panel C indicates the performance of other targeting methods the government could have used. Accuracy, precision, and recall are evaluated by the extent to which they reach the poorest 29% (by construction, precision and recall are equal in this simulation). Standard deviations, produced from 1,000 bootstrap simulations, shown in parentheses.

Exclusion Source	Proportion Included	Data and Calculations
Voter ID possession	83% - 98%	According to administrative data, 3,633,898 individuals are registered to vote in Togo. The electoral commission of Togo reports that this corresponds to 86.6% of eligible adults[xvii]. The total adult population in Togo is not certain (the last census was in 2011), but Togo's national statistical agency (https://inseed.tg/) estimates that there are 3,715,318 adults in Togo; the United Nations estimates 4.4 million adults[47]. These imply a voter ID penetration rate of either 82.6% and 97.8%, respectively
SIM card and mobile phone access	65% - 85%	65% of individuals interviewed in the 2018-2019 field survey (N = 6,171) report owning a mobile phone; 85% of individuals are in a household with one or more phones. See Figure S9. It is likely that phone penetration in Togo increased between the time of the field survey and the Novissi expansion in mid-2020; the Togolese government estimates 82% SIM card penetration in the country (though some people may have multiple SIM cards)[xviii].
Program awareness	35% - 46%	245,454 unique subscribers attempted to register for the rural Novissi program. The total voting population of eligible areas is 528,562, implying a maximum registration rate of 46.44%. However, not all 245,454 registration attempts were made by people living in eligible areas; examining administrative data on home location from successful registrations we estimate that 87% of registration attempts came from eligible areas, implying an attempted registration rate of 40.40%. An alternative way to estimate attempted registration rates involves comparing the number of registration attempts made by phones below the poverty threshold (69,753) to our estimate of the number of voters in eligible cantons below the poverty threshold based on inferred home locations from mobile phone data (174,425, see Appendix II for details), which implies an attempted registration rate of 34.79% after scaling by 87% (to account for registrations that came from outside of eligible areas).
Registration challenges	72%	Registration for the Novissi program requires entering basic information into a USSD (phone-based) platform. According to program administrative data, of the 245,454 subscribers who attempted registration, 176,517 (71.95%) eventually succeeded. The average registration required four attempts.
Targeting errors	47%	Based on the estimates from our targeting simulations using the 2020 phone survey (Table 1), the exclusion error rate of the phone-based targeting algorithm is 53%.

Table 2 | Sources of exclusion from rural Novissi benefits. We use multiple sources of administrative data, survey data, and government sources to estimate the extent to which different elements of the Novissi program design may have led to errors of exclusion. Novissi eligibility requirements include: voter ID possession (to provide identification and home location), access to a mobile phone (to fill register using the USSD platform), program awareness (to know that the program exists and to attempt to register), ability to register via the USSD platform (which requires basic digital literacy), as well as targeting errors from the phone-based machine learning algorithm. While this table calculates sources of exclusion as though they were all independent, Table S8 uses survey data to calculate overlaps in exclusions.

[xvii] https://www.republicoftogo.com/Toutes-les-rubriques/Politique/86-6-des-Togolais-ont-une-carte-d-electeur
[xviii] Ibid.

Supplementary Information

Appendix A: Selection of Variables for Proxy-Means Test

Our proxy-means test is used in analysis for both the 2018-2019 field survey (where we evaluate the PMT's accuracy as a targeting mechanism) and the 2020 phone survey (where we use the PMT as a measure of ground-truth poverty in the absence of a consumption measure). We construct the PMT using all observations from the 2018-2019 field survey (N=6,171). We begin by identifying all information on demographics and asset ownership collected in the field survey that may correlated with poverty. In total, we identify 56 variables, including information on household assets and housing quality, education, marital status, age, ethnicity, health, location, and more.

Our goal is to identify a small subset of variables that are most predictive of household consumption. We use stepwise forward selection of predictors to greedily identify the most predictive feature subsets of size K, for K ranging from 1 to 30. Specifically, we randomly divide our survey observations into a training set (75%) and test set (25%). For K=1, we train a machine learning model to predict household consumption[xix] from each feature individually, and select the feature associated with the best model. For $K=2$, we test adding each remaining feature to our model, and select the feature that adds the most predictive power. We continue the process for all K up to 30.

We perform the stepwise forward selection process first for a Ridge regression (where the optimal L2 penalty is selected via awide grid) and second for a random forest (where the optimal ensemble size is chosen from {50, 100} via 3-fold cross validation and the optimal tree depth is chosen from {2, 4, 6, 8}). Figure S10 plots the predictive accuracy (measured with r^2 score) for each value of K for the two models.

We observe that the random forest is not significantly more accurate than the regression, and note a greater degree of overfitting with the random forest. We therefore select the Ridge regression, as it has the advantages of simplicity and interpretability without sacrificing accuracy. We identify an "elbow" in the accuracy progression at K=11 features, so we use the feature subset of size $K=11$ in our PMT. These features and the weights associated with them are recorded in Table S2.

[xix] While in the rest of this paper we use price-index adjusted per capita household consumption, in this exercise our outcome variable is raw household consumption (because the data necessary to construct price index adjusted consumption was not available to us at the time the exercise was carried out prior to the 2020 phone survey).

Appendix B: Sampling for 2020 Phone Survey

The 2020 phone survey was sampled to be as representative as possible of mobile phone subscribers living in the cantons of Togo eligible for rural Novissi. The sample frame for the survey is all mobile phone subscribers active on one of the two mobile networks in Togo between March 1 and September 30, 2020 (N=5.83 million). Sampling is based on four metrics associated with each mobile phone subscriber: inferred probability of living in a rural Novissi-eligible area, registration to a previous Novissi program, inferred wealth based on our CDR+ML methods, and total mobile phone expenditure.

Inferred probability of living in a rural-Novissi eligible area

We use machine learning methods to predict the probability that each of the 5.83 mobile network subscribers is living in one of the 100 cantons of Togo that are eligible for rural Novissi. We train the machine learning model on the dataset of all subscribers that registered for Novissi when it was first available in the Greater Lomé region (while only residents of Greater Lomé were eligible for this program, any registered voter in Togo could sign up for the platform for immediate eligibility in future programs). In total, this dataset includes 1.3 million subscribers with Novissi registration data matched to CDR. Novissi registration data crucially includes the canton in which each subscriber registered to vote (we refer to this information as ground-truth home cantons).

The raw training dataset is not representative of mobile network in Togo, as a nonrandom subset of subscribers register for Novissi (for example, more than half of the registered subscribers are in the Greater Lomé region). To make the training data more representative, we calculate the expected proportion of subscribers in each canton based on the total number of voters registered in each canton and individual mobile phone penetration measured at the prefecture level (based on the 2018-2019 field survey). We "balance" the training dataset by sampling observations at random from cantons with a disproportionately high number of registrants until the proportions in the training dataset reflect the expected proportion of mobile network subscribers in each canton.

Finally, we train a machine learning model to predict whether each subscriber is in the 100 eligible cantons. As in poverty prediction, we use a gradient boosting model with optimal hyperparameters chosen via cross-validation. The model uses CDR features relating to location – specifically the (normalized) number of unique days on which each subscriber places a transaction in each canton of Togo. The model obtains an AUC score of 0.90 over cross-validation, and an overall accuracy of 93%. We use the machine learning model to produce estimates of the likelihood that all 5.83 million mobile network subscribers live in an eligible canton.

Registration to a previous Novissi program

As noted above, around 22% of mobile network subscribers were already registered in the Novissi system prior to rural Novissi, and therefore are associated with a ground-truth home canton based on the canton in which they are registered to vote. In our dataset of inferred home location likelihoods, we assign any subscriber registered to vote in one of the 100 targeted cantons a 100% likelihood of geographical eligibility (N=86,856). We assign any subscriber registered to vote outside of these cantons a 0% likelihood of geographical eligibility (N=1,046,905).

Based on total number of voters registered in targeted cantons and mobile phone penetration in these areas, we estimate that around 240,000 subscribers live in eligible cantons. We identify the 240,000 subscribers most likely to be living in a targeted canton (including all 86,856 subscribers registered in targeted cantons). Only these 240,000 subscribers are eligible to be surveyed.

Inferred poverty based on CDR

We use ground-truth poverty data collected in a previous nationally-representative phone survey conducted in June 2020 to train a machine learning model to predict poverty from CDR. We follow the methods described in Methods §15, using the PMT as ground truth and CDR features from March 1 to September 30, 2020. We use the machine learning model to predict the poverty of each of the 5.83 million mobile phone subscribers on the networks.

Mobile phone expenditure

We construct the measure of total phone expenditure for each subscriber described in Methods §4.

Survey sampling

We seek to maximize power around the targeting thresholds by oversampling observations close to the targeting thresholds. We divide the 240,000 subscribers into four quartiles based on inferred poverty based on CDR and mobile phone expenditure. We overlap the quartiles to form eight "cells", based on the combination of the two targeting methods (for example, cell AA represents being in the lowest quartile by both targeting methods, while cell AD represents being in the lowest quartile by one method and the highest quartile by another method, and so on). We assign a weight of 0.20 to cells AD and BC, a weight of 0.15 to cells AC and BD, a weight of 0.10 to cells AB and AA, and a weight of 0.05 to cells CD and DD. We sample from our dataset of 240,000 subscribers with the associated probability. We use the inverse of these sampling probabilities as sample weights in our downstream analysis, in combination with response weights (see Methods §10).

Supplementary Figures

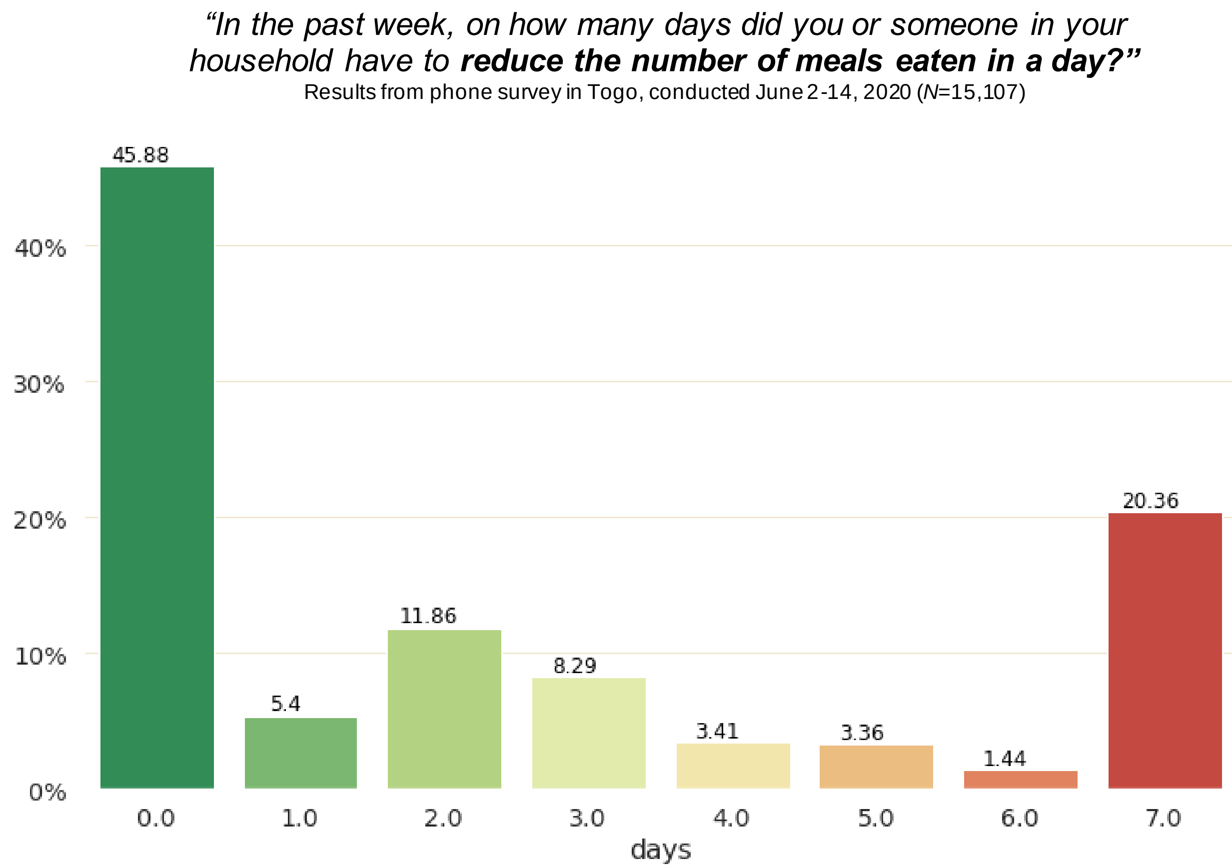

Figure S1 | Food insecurity in Togo. In June 2020, we conducted a phone survey of 15,107 mobile phone owners in Togo. Survey weights are used to make responses representative of the population of mobile phone owners in Togo.

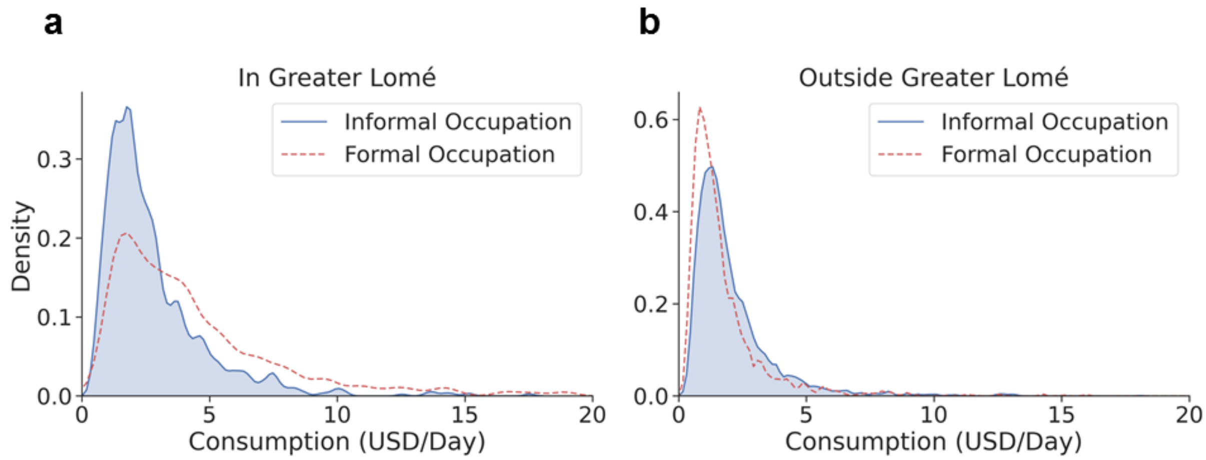

Figure S2 | Wealth of formal vs. informal workers. Results based on analysis of nationally-representative household survey data collected by the Government of Togo in 2018-2019 (N = 6,171). Data is collected at the household-level, we assign a household-level informal occupation indicator if at least one of the adult household members is unemployed or employed in an informal occupation. See Methods §3.

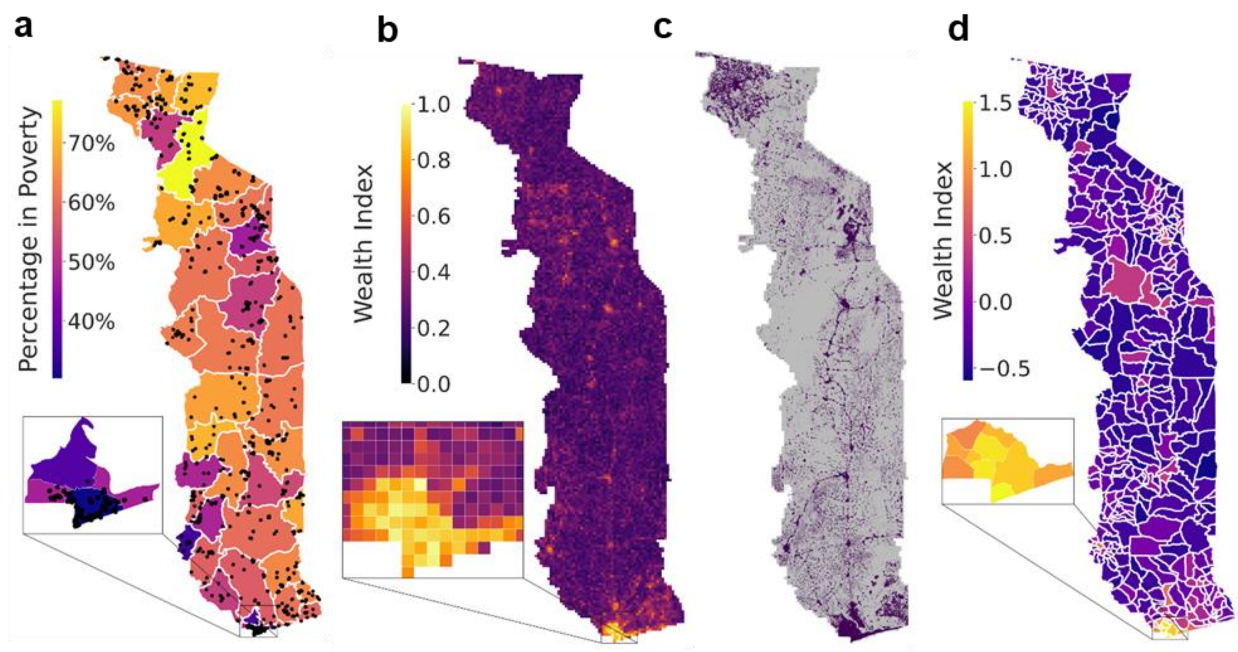

Figure S3 | Poverty maps. (a) Prefecture (admin-2) poverty map inferred from 2017 field survey (N = 26,902), showing the percent of the population living under the poverty line by prefecture. Overlayed with locations of survey observations in black points. **(b)** High-resolution estimates of consumption derived from satellite imagery. **(c)** High-resolution estimates of population density derived from satellite imagery. **(d)** Canton (admin-3) poverty map inferred from satellite imagery by combining high-resolution consumption estimates and population density estimates to calculate weighted average consumption per canton.

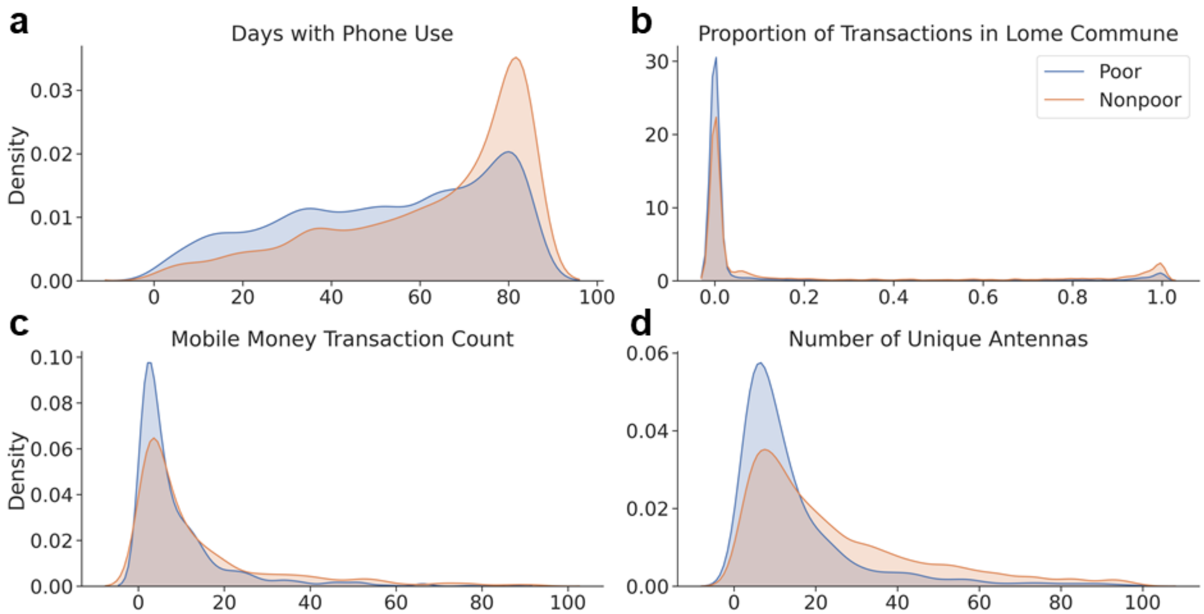

Figure S4 | CDR features. Comparing the distribution for CDR features for those above and below the international poverty line (USD 1.90/day) in the 2018-2019 field survey dataset.

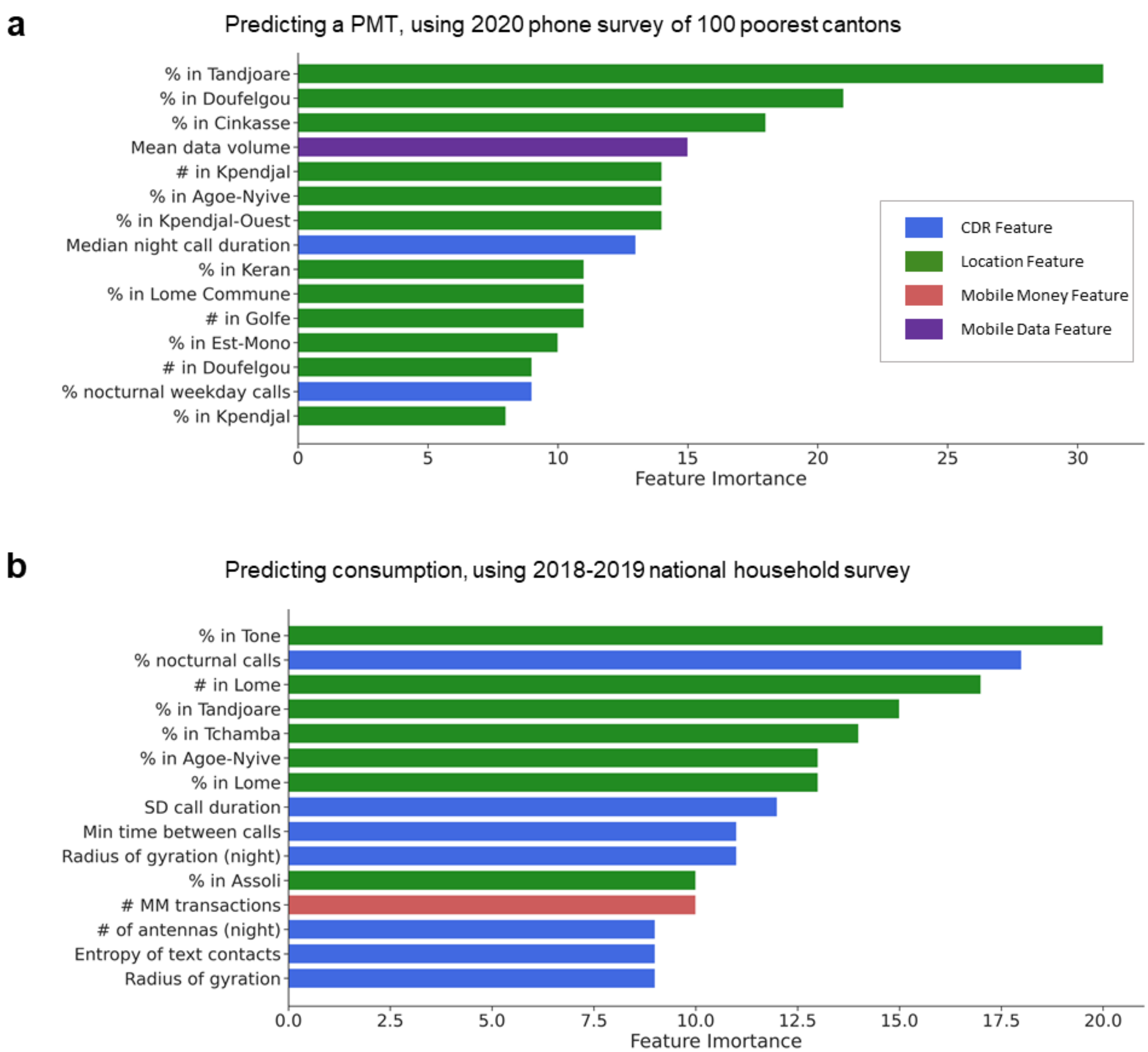

Figure S5 | Feature importances for ML models. Feature importances for the machine learning models trained to predict **(a)** Proxy Means Test from CDR, using a 2020 phone survey of mobile subscribers in Togo's 100 poorest cantons ($N = 8{,}915$); and **(b)** consumption from CDR in the 2018-2019 field survey dataset ($N = 4{,}171$). Feature importance is calculated based on the total number of times a feature is split upon in the prediction ensemble. Features are color-coded as follows: CDR features are shown in blue, location features in green, mobile money features in red, and mobile data features in purple.

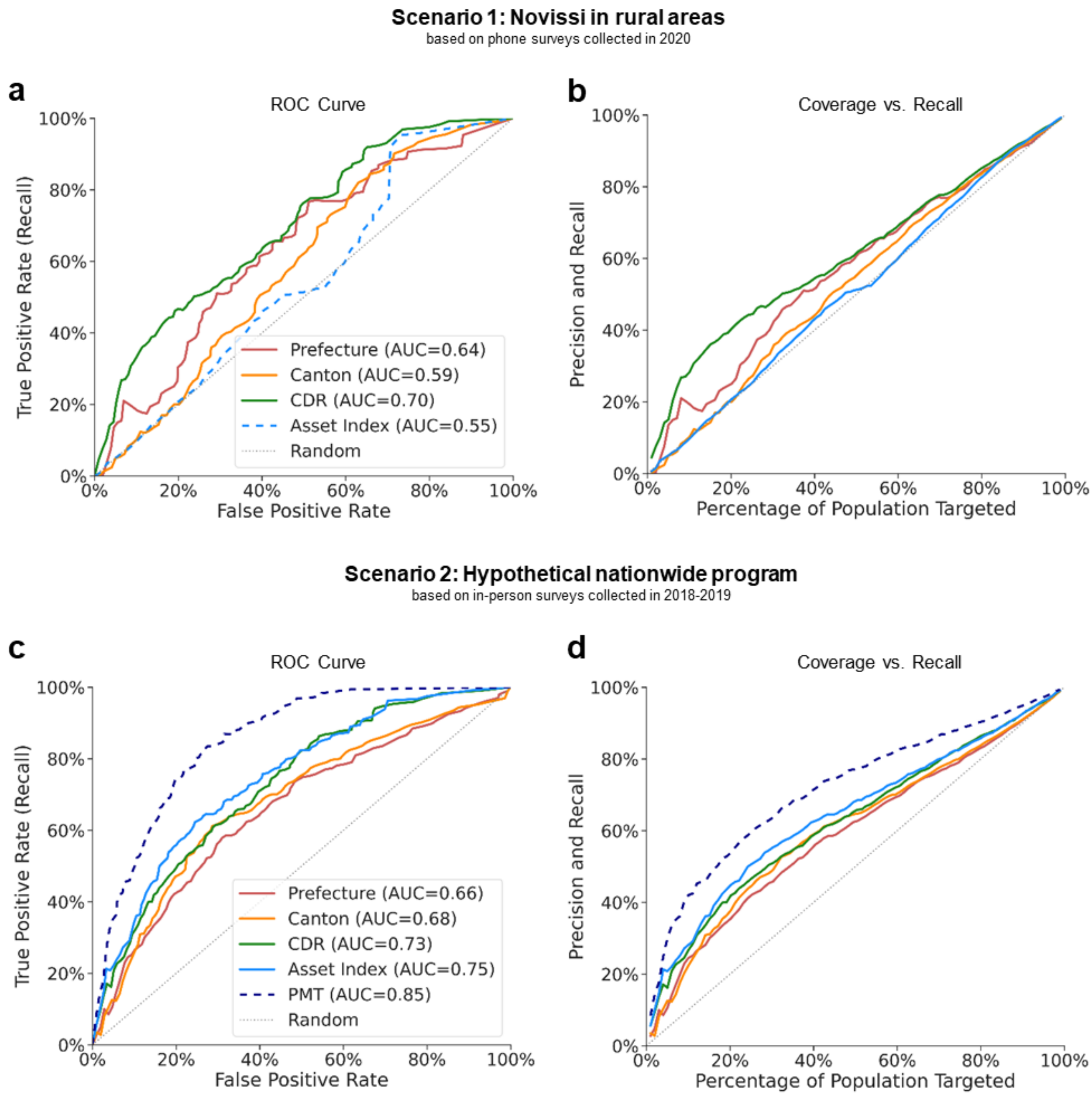

Figure S6 | Targeting performance at different levels of program coverage. Top figures (**a** and **b**) show performance for the rural Novissi program, evaluated using 2020 phone survey. Bottom figures (**c** and **d**) correspond to the hypothetical national program, evaluated using the 2018-2019 field survey. ROC curves on left (**a** and **c**) indicate the true positive and false positive rates at different targeting thresholds, as described in Methods §5.c. Coverage vs. Recall figures on right (**b** and **d**) show how precision and recall vary as the percentage of the population receiving benefits increases, i.e., they indicate the precision and recall for reaching the poorest *k%* of the population in programs that target the poorest *k%*. (Precision and recall are thus the same for each value of *k* by construction – see Methods §5.c).

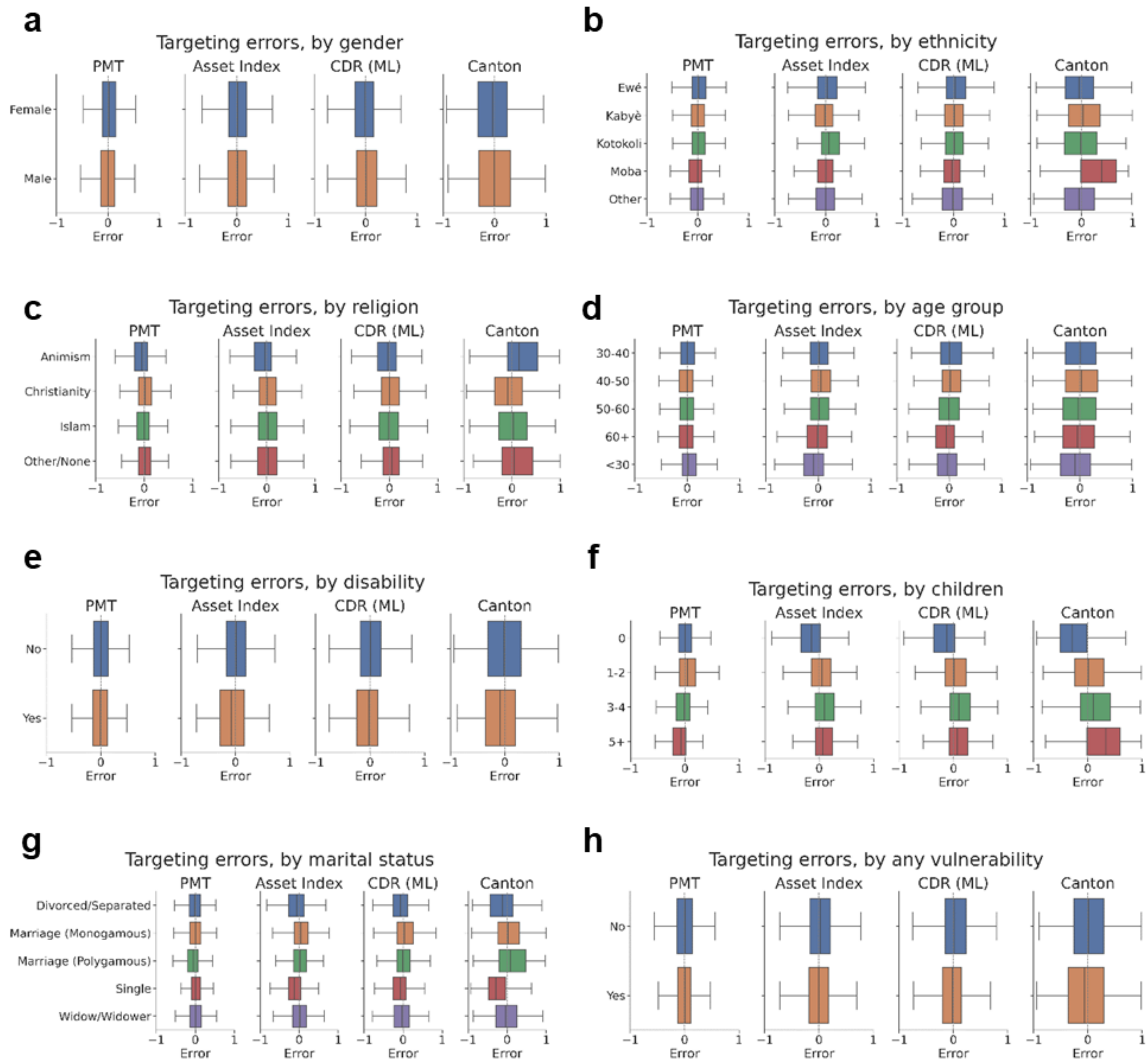

Figure S7 | Fairness with normalized rank residuals. Boxplots showing distributions of normalized rank residuals (see Methods §5.e) aggregated by subgroup, using the 2018-2019 field survey dataset (N = 4,171). Left-shifted boxes indicate groups that are consistently under-ranked by a given targeting mechanism, right-shifted boxes indicate groups that are consistently over-ranked by a given targeting mechanism.

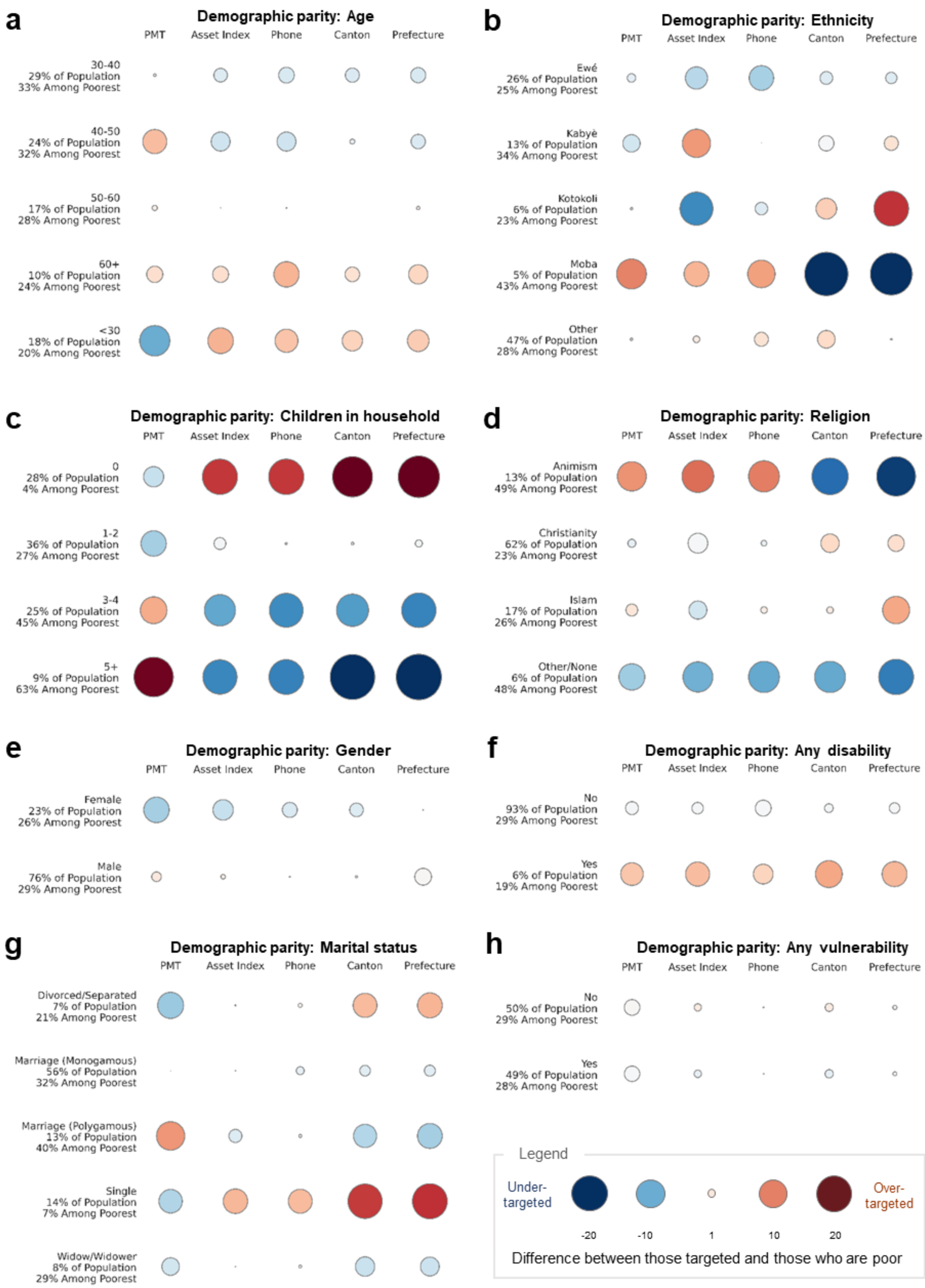

Figure S8 | Fairness with demographic parity. We evaluate demographic parity across subgroups by comparing the proportion of a subgroup targeted under counterfactual approaches to the proportion of the subgroup that falls into the poorest 29% of the population (using data from the 2018-2019 field survey matched to CDR, N = 4,171). Bubbles show the percentage point difference between the proportion of the subgroup that is targeted and the proportion that is poor according to ground-truth data. Large red bubbles indicate groups that are over-targeted; large blue bubbles indicate groups that are under-targeted.

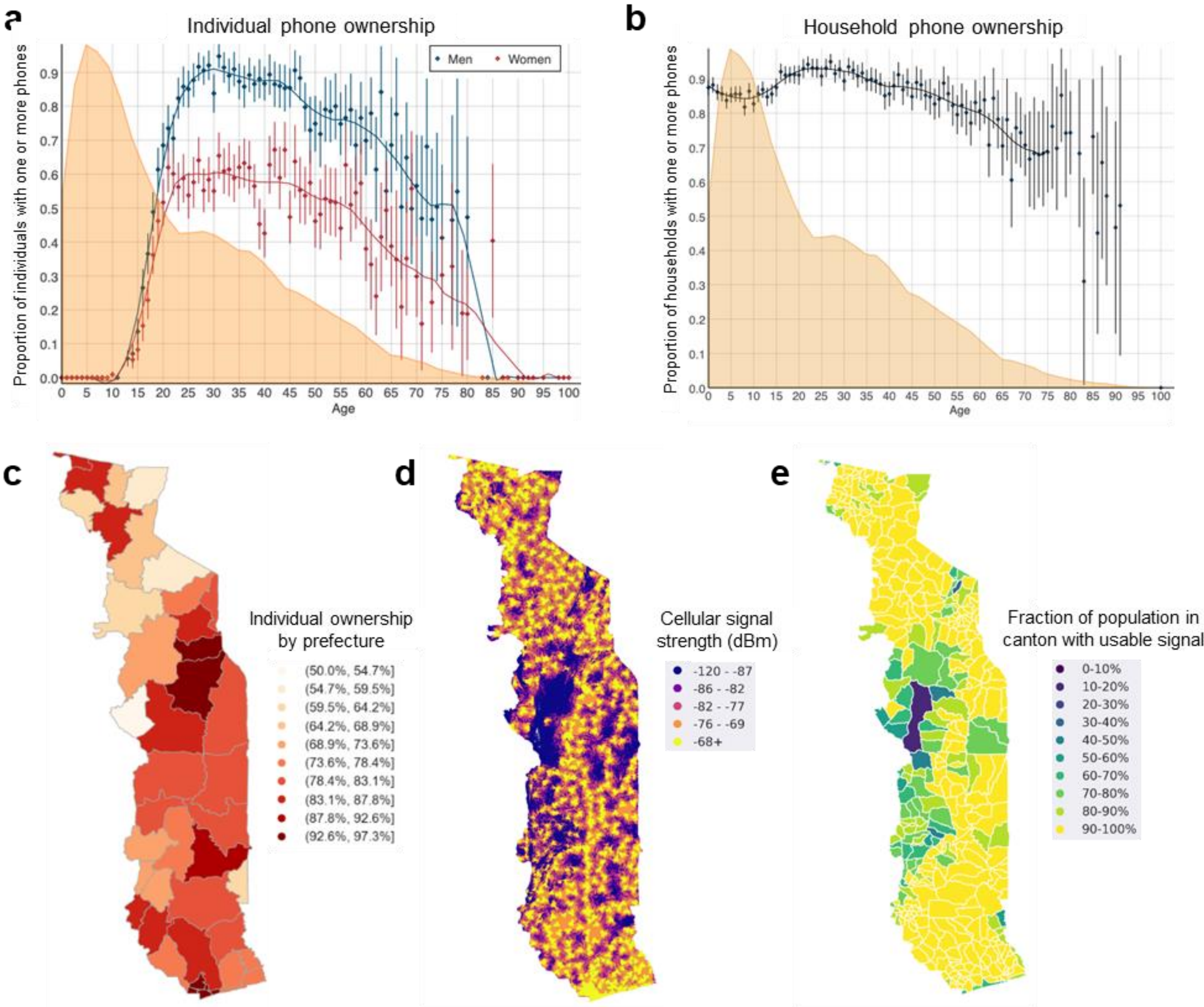

Figure S9 | Mobile phone penetration and coverage in Togo. Based on nationally-representative household survey data collected in 2018-2019, we estimate **a)** the percentage of adults in Togo with one or more mobile phone, disaggregated by age and gender (vertical bands indicate 95% confidence intervals); **b)** the percentage of households in Togo with one or more mobile phones, disaggregated by the age of the head of household; and **c)** the percentage of individuals in each prefecture with one or more mobile phones. Using data on the location and signal strength of all cell towers in Togo, made available by Togocel (one of the two phone companies in Togo), we calculate **d)** the signal strength across Togo; and **e)** the fraction of the population in each canton with access to a usable signal, where signal greater than -86 dBm is generally considered usable, and sub-canton estimates of population density are derived from satellite imagery and downloaded from the Humanitarian Data Exchange[41].

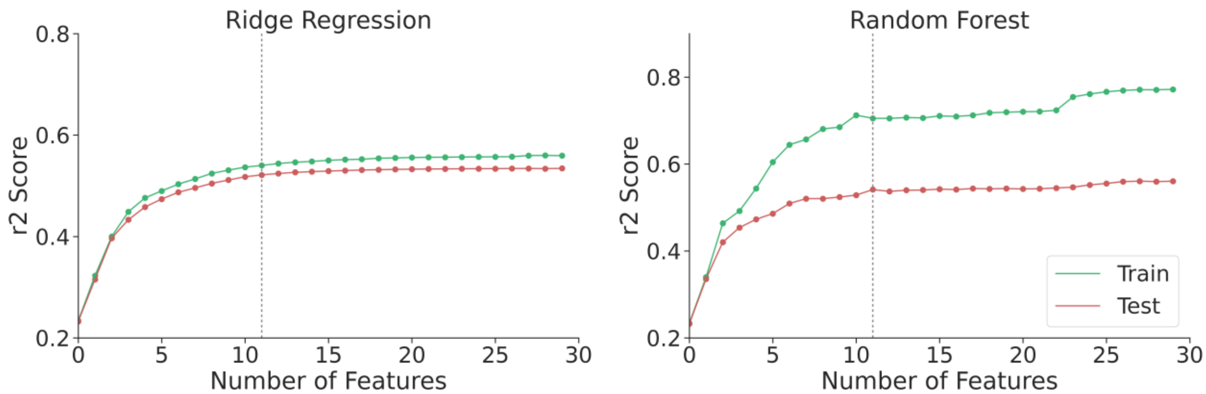

Figure S10 | Selection of variables for proxy-means test. Each plot shows the accuracy (measured by r^2 score) of a proxy-means test using the most predictive feature subset of size *K*, where *K* is plotted on the x-axis. The left plot shows the accuracy obtained by a Ridge regression; the right plot shows the accuracy obtained by a random forest. Feature subsets are selected via stepwise forward selection.

Supplementary Tables

Asset	Magnitude (2018-2019 Field Survey)	Magnitude (2020 Phone Survey)
Electricity access	0.38	
Toilet	0.37	0.41
TV	0.35	
Electricity grid	0.35	
Garbage disposal	0.33	
Waste disposal	0.33	
Iron	0.26	0.06
Radio	0.20	0.23
Clean water (wet season)	0.16	
Clean water (dry season)	0.16	
Refrigerator	0.12	0.02
Walls	0.12	
Floor	0.11	
Mobile phone	0.11	
Water disposal	010	
Motorcycle	0.10	0.88
Computer	0.09	0.02
Roof	0.08	
Stove	0.07	0.06
Car	0.06	0.00
Tablet	0.01	0.00
Air conditioner	0.01	0.00
House	0.00	
Electricity (offgrid)	0.00	

Table S1 | Asset-based wealth index. Magnitude of first principal component for 2018-2019 field survey and 2020 phone survey.

Feature	β	Feature (continued)	β
Car	2.77	HHW Education 4	-0.18
Stove	1.77	Pref. Lacs	-0.18
Refrigerator	1.32	Pref. Sotouboua	-0.18
HHH Education 8	1.12	Pref. Kloto	-0.21
HHH Education 9	0.91	HHW Education 6	-0.21
HHH Hospitalization	0.81	Pref. Kpele	-0.21
Iron	0.63	Pref. Bas-Mono	-0.23
HHH Education 3	0.55	Pref. Lome Commune	-0.23
TV	0.50	Pref. Danyi	-0.24
All children in school	0.48	Pref. Yoto	-0.26
Pref. Cinkasse	0.39	Pref. Agoe-Nyive	-0.27
Pref. Tchamba	0.33	HHH Education 5	-0.27
Toilet	0.26	No children in school	-0.31
HHH Education 7	0.17	Pref. Assoli	-0.32
Pref. Est-Mono	0.14	Pref. Kpendjal-Ouest	-0.33
HHW Education 0	0.12	Pref. Zio	-0.33
Pref. Tchaoudjo	0.09	Pref. Amou	-0.34
Pref. Bassar	0.09	HHW Education 3	-0.34
Pref. Haho	0.07	Pref. Plaine du Mo	-0.34
Pref. Dankpen	0.04	Pref. Anie	-0.34
Pref. Moyen-Mono	-0.03	Pref. Tandjoare	-0.35
Pref. Oti-Sud	-0.06	Pref. Binah	-0.37
Pref. Oti	-0.08	Pref. Ave	-0.39
Pref. Wawa	-0.11	Pref. Keran	-0.41
Pref. Vo	-0.11	Pref. Kpendjal	-0.46
Pref. Ogou	-0.12	HHW Education 2	-0.50
Pref. Tone	-0.14	Pref. Kozah	-0.51
Pref. Agou	-0.15	HHH Education 2	-0.57
Pref. Akebou	-0.17	Pref. Blitta	-0.61
HHW Education 1	-0.17	HHH Education 1	-0.63
Some children in school	-0.17	Pref. Golfe	-0.68
Number of children	-0.17	Pref. Doufelgou	-0.75

Table S2 | Proxy means test. Weights for linear model, trained on 2018-2019 phone survey (*N* = 6,171).

	2018-2019 Field Survey (N=6,171)			2020 Phone Survey (N=8,915)	
	Consumption	Proportion	N	Proportion	N
Intellectual Professions	$4.11 (3.55)	7%	277	7%	577
Intermediate Professions	$3.95 (3.56)	5%	197	3%	264
Administrators	$3.89 (3.57)	1%	32	0%	16
Managers and Directors	$3.70 (3.03)	3%	106	0%	36
Unemployed/Unknown	$3.19 (2.44)	8%	339	3%	275
Direct Services and Merchants	$2.75 (2.11)	23%	940	28%	2,111
Industry/Artisans	$2.47 (1.83)	15%	587	12%	1,026
Military Professions	$2.45 (1.25)	0%	17	1%	26
Elementary Professions	$2.21 (1.83)	2%	65	3%	249
Factory Workers	$2.17 (1.44)	7%	267	2%	165
Agricultural Professions	$1.53 (0.94)	29%	1,744	41%	4,170

Table S3 | Occupation categories. Average daily per capita consumption per occupation category, with counts by category, separately for the 2018-2019 field survey and 2020 phone survey. Occupation categories for the 2018-2019 survey are for the household head, for the 2020 survey are for the individual respondent.

	Consumption	PMT	Asset Index
Panel A: 2018-2019 Field Survey (N = 4,171)			
ML	0.46	0.62	0.74
Single Feature	0.13	0.16	0.11
Panel B: 2018-2019 Field Survey, Rural Only (N = 2,306)			
ML	0.31	0.44	0.51
Single Feature	0.09	0.12	0.08
Panel C: 2020 Phone Survey (N = 8,915)			
ML	--	0.41	0.40
Single Feature	--	0.13	0.14

Table S4 | Performance of phone-based approach to predicting wealth and consumption. Accuracy (Pearson correlation coefficients) for predicting poverty measures from CDR. ML predictions are produced over 5-fold cross validation and evaluated for pooled correlation. The "single feature" model estimates wealth and consumption based on the individual's total expenditures on calling and texting.

	Targeting Novissi in rural Togo Based on 2020 Phone Survey (N = 8,915)			Hypothetical nationwide program Based on 2018-2019 Field Survey (N = 4,171)		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Panel A: Targeting methods considered by the Government of Togo in 2020						
Prefecture (Admin-2 regions)	47% (0.67%)	86% (1.16%)	34% (0.46%)	60% (0.67%)	68% (1.15%)	39% (0.67%)
Canton (Admin-3 regions)	44% (0.87%)	80% (1.51%)	31% (0.59%)	62% (0.69%)	71% (1.19%)	41% (0.68%)
Phone (Expenditures)	41% (0.77%)	76% (1.32%)	30% (0.52%)	57% (0.91%)	63% (1.56%)	36% (0.90%)
Phone (Machine Learning)	48% (0.76%)	87% (1.30%)	34% (0.51%)	63% (0.69%)	72% (1.19%)	42% (0.69%)
Panel B: Common alternative targeting methods that could not be implemented in Togo in 2020						
Asset Index	42% (0.52%)	77% (0.89%)	30% (0.35%)	65% (0.69%)	76% (1.19%)	44% (0.68%)
PPI		*[data not available]*		69% (0.66%)	83% (1.14%)	48% (0.66%)
PMT		*[data not available]*		71% (0.56%)	87% (0.97%)	50% (0.56%)
Panel C: Additional counterfactual targeting methods that were feasible in Togo in 2020						
Random	39% (0.76%)	73% (1.31%)	29% (0.51%)	49% (0.88%)	49% (1.51%)	28% (0.87%)
Occupation (As implemented)	38% (0.77%)	71% (1.33%)	28% (0.52%)	48% (0.61%)	46% (1.05%)	27% (0.60%)
Occupation (Optimally designed)	46% (0.61%)	84% (1.06%)	33% (0.42%)	64% (0.68%)	74% (1.18%)	43% (0.68%)

Table S5 | Performance of targeting households below the poverty line. Analysis is similar to that presented in Table 1, but targeting is evaluated on the extent to which each method (still targeting the poorest 29%) provides benefits to individuals consuming less than the international poverty line of USD $1.90 per person per day (76% of observations in the 2020 phone survey dataset and 57% of observations in the 2018-2019 field survey). Spearman correlation and AUC are not reported here as they do not depend on the classification threshold, and are thus identical to the values reported in Table 1.

	Targeting Novissi in rural Togo Based on 2020 Phone Survey (N = 8,915)			Hypothetical nationwide program Based on 2018-2019 Field Survey (N = 4,171)		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Panel A: Targeting methods considered by the Government of Togo in 2020						
Prefecture (Admin-2 regions)	59% (0.94%)	61% (1.49%)	37% (0.99%)	67% (0.73%)	51% (1.26%)	44% (1.09%)
Canton (Admin-3 regions)	54% (0.86%)	53% (1.47%)	32% (0.91%)	69% (0.73%)	54% (1.26%)	47% (1.08%)
Phone (Expenditures)	53% (0.85%)	50% (1.32%)	31% (0.90%)	64% (0.85%)	45% (1.46%)	39% (1.25%)
Phone (Machine Learning)	61% (0.77%)	64% (0.94%)	39% (0.81%)	69% (0.73%)	55% (1.27%)	48% (1.09%)
Panel B: Common alternative targeting methods that could not be implemented in Togo in 2020						
Asset Index	53% (0.54%)	51% (0.009)	31% (0.57%)	72% (0.71%)	60% (1.23%)	51% (1.05%)
PPI		*[data not available]*		76% (0.73%)	67% (1.26%)	57% (1.09%)
PMT		*[data not available]*		78% (0.70%)	70% (1.20%)	60% (1.03%)
Panel C: Additional counterfactual targeting methods that were feasible in Togo in 2020						
Random	53% (0.84%)	51% (1.31%)	31% (0.88%)	56% (0.81%)	33% (1.39%)	28% (1.20%)
Occupation (As implemented)	47% (0.76%)	41% (1.17%)	25% (0.80%)	54% (0.55%)	29% (0.96%)	25% (0.82%)
Occupation (Optimally designed)	59% (0.68%)	61% (1.61%)	37% (0.71%)	71% (0.74%)	58% (1.28%)	50% (1.10%)

Table S6 | Performance of targeting households below the *extreme* poverty line. Analysis is similar to that presented in Table 1, but targeting is evaluated on the extent to which each method (still targeting the poorest 29%) provides benefits to individuals consuming less than the international *extreme* poverty line, set at 75% of the international poverty line or USD $1.43 per person per day (53% of observations in the 2020 phone survey dataset and 41% of observations in the 2018-2019 field survey). Spearman correlation and AUC are not reported here as they do not depend on the classification threshold, and are thus identical to the values reported in Table 1.

Targeting hypothetical nationwide program – but only in rural areas
Based on 2018 Phone Survey Restricted to Rural Areas (N = 2,306)

	Spearman	AUC	Accuracy	Precision & Recall
Panel A: Targeting methods that were feasible in Togo in 2020				
Prefecture (Admin-2 regions)	0.16 (0.023)	0.57 (0.011)	64% (0.97%)	37% (1.67%)
Canton (Admin-3 regions)	0.19 (0.025)	0.59 (0.013)	63% (0.98%)	36% (1.69%)
Phone (Expenditures)	0.15 (0.024)	0.59 (0.012)	63% (1.05%)	36% (1.81%)
Phone (Machine Learning)	0.30 (0.023)	0.65 (0.012)	67% (1.00%)	43% (1.73%)
Panel B: Common alternative targeting methods that could not be implemented in Togo in 2020				
Asset Index	0.36 (0.023)	0.68 (0.011)	67% (1.01%)	44% (1.74%)
PPI	0.55 (0.017)	0.77 (0.009)	72% (1.07%)	52% (1.84%)
PMT	0.61 (0.016)	0.80 (0.007)	73% (1.06%)	54% (1.84%)
Panel C: Additional counterfactual targeting methods that were feasible in Togo in 2020				
Random	0.00 (0.024)	0.50 (0.012)	59% (1.04%)	29% (1.79%%)
Occupation (Novissi)	-0.13 (0.024)	0.44 (0.011)	54% (0.89%)	21% (1.53%)
Occupation (Optimal)	0.31 (0.023)	0.63 (0.010)	63% (0.62%)	37% (1.06%)

Table S7 | Performance of targeting the hypothetical national program, when restricted to rural areas. Analysis is similar to that presented in the last four columns of Table 1, but analysis is restricted to the 2,306 survey respondents (of the 4,171 total) who live in rural areas.

Exclusion Source	N	Succeed	Drop Out	% of Total Remaining
All poor	3,172	--	--	100.00%
Own a voter ID	3,170	99.77%	0.23%	99.77%
Attempt to register for Novissi	2,230	60.56%	39.44%	60.43%
Succeed in registration	1,826	78.68%	21.32%	47.54%
Targeted by phone PMT	1,242	60.43%	39.57%	28.73%

Table S8 | Overlapping sources of exclusion from rural Novissi. Progressive sources of attrition from the rural Novissi program, where each row shows exclusion conditional on exclusions from preceding rows. Results are estimated using the 2020 phone survey ($N = 8{,}915$). Values reweighted using sample weights.