Machine Learning and Mobile Phone Data Can Improve Aid Targeting in Low- and Middle-Income Countries

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 around the world have mobilized targeted social assistance programs[2]. A central challenge in the administration of these programs is *targeting*: how to identify the individuals and families with the greatest need[3]? This challenge is particularly acute in the large number 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 to 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 anti-poverty 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 other targeting options available to the Government of Togo at the time, the machine learning approach reduced errors of exclusion by between 2% and 50%. 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 4-30%. 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to contribute to humanitarian response efforts, particularly in crisis settings when traditional data are missing or out of date.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) estimate that over 1,400 new social assistance programs have been launched since early 2020[2], providing over $800 billion dollars in cash transfer payments to over 1.7 billion people (almost 17% 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 were used to determine 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 LMICs, where economic activity is often informal and based on home-produced agriculture, the government typically does 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 be used to 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 social assistance program. The Government of Togo launched Novissi in March 2020, shortly after the first COVID-19 cases appeared. 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 face-to-face 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, they did not have a traditional social registry that could be used to determine program eligibility, and they had neither the time nor the resources to

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

build 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 disbursed to individuals who met three criteria: (1) "self-targeted"[15] by dialing in to the Novissi platform from a mobile phone; (2) registered to vote in specific regions (the program initially focused on the Greater Lomé region around the capital city); and (3) self-declared an informal occupation.

Targeting informal occupations may have been an effective strategy to prioritize benefits to people who were forced to stop working at the onset of the crisis, this approach does not necessarily target benefits to the poorest households in the country. Based on the analysis of nationally-representative household survey data collected prior to the onset of COVID-19, we find that while informal workers in Greater Lomé were poorer than formal workers (Figure S2a), in rural areas occupation-based targeting would actually provide a greater share of benefits to wealthier households (Figure S2b and Methods §5).

To assist the government in directing benefits to the poorest individuals in the country, we developed a two-step approach to targeting aid based on high-resolution satellite imagery and passively collected mobile phone metadata (Figure 1). The two steps were designed to meet the government's two stated policy objectives: first, to concentrate benefits in the poorest geographic regions of the country; and second, to prioritize benefits to the poorest mobile subscribers in those regions.

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 wealth 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, with two main differences:

[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 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 small number of questions about the assets and characteristics of the household. To construct a national registry usually requires that government surveyors visit all households in the country, and typically takes years to organize and execute[19].

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 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, by comparing targeting errors using the phone-based PMT to targeting errors under four different counterfactual approaches. We present results from two scenarios: First, the actual policy scenario faced by the Government of Togo in September of 2020, which required distributing cash to 60,000 beneficiaries within Togo's 100 poorest *cantons* (the smallest administrative unit in the country, which contains on average 10 villages); this first scenario is evaluated using data collected in a large phone survey we conducted in September 2020, which was designed for this purpose. Second, we simulate and evaluate a more general policy scenario in which the government is not limited to specific geographies; this scenario is evaluated using national household survey data collected in the field by the government in 2018-2019. These data are described in Methods §3 and details on the evaluation are in Methods §5.

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 mistakenly deemed eligible), relative to the other feasible approaches to targeting available to the Government of Togo (Table 1, Panel A). We focus on the ability of each targeting method to reach the poorest 29% in each of the two survey datasets, since rural Novissi benefits were provided to 29% of program registrants. In the first scenario focused on reaching the poorest people in the 100 poorest cantons, using a PMT as a measure of "true" poverty status (since the phone survey did not collect consumption data), we find that phone-based targeting (AUC= 0.69) outperforms the other feasible methods of targeting rural Novissi aid (e.g., AUC=0.57-0.58 for geographic blanket targeting). As a result, errors of exclusion (defined as 1 – Recall) are lower for the phone-based approach (55%) than for feasible alternatives (56%-78%).

Phone-based targeting is likewise preferred to the other three feasible methods when we simulate the targeting of a national anti-poverty program (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 = 0.64 - 0.67), and likewise leads to lower exclusion errors (49%) than all feasible alternatives (53%-73%). Note that exclusion errors for *all* evaluated methods are quite high, given the difficulty of identifying the poorest 29% of a population. Table S4 and Table S5 present targeting results for alternative targeting thresholds, to measure errors at targeting households below the poverty line and extreme poverty line.

For completeness, we also compare the phone-based approach to common alternative approaches that require a recent and comprehensive social registry (Table 1, Panel B). 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. Here, the results are mixed: the phone-based approach (AUC = 0.73) is approximately as accurate as targeting using an asset-based wealth index (AUC = 0.75), but less accurate than using a perfectly calibrated PMT (AUC = 0.85). 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.a)[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 function of consumption and the benefits assigned to an individual. 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. Note that this social welfare analysis also allows us to identify the optimal beneficiary share and corresponding transfer size. Figure 2 shows the utility curves for each of the targeting methods simulated, separately for the two datasets. Note that CDR-based targeting, geographic blanketing, and an asset-based wealth index all achieve approximately the same maximum utility in the 2018-2019 field survey dataset, but CDR-based targeting dominates in the 2020 phone survey dataset.

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[24,25]. 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, though we note that notions of fairness and parity are in tension and contested.[26,27] Figure 3a shows, as an example, that the ML approach does not cause women to be systematically more likely to be incorrectly excluded from receiving benefits than men (see also Methods §5.d). Likewise, the ML approach does not create systematic exclusion errors for specific ethnic groups (Figure 3b), religions, or age groups (Figure S8). We also compare the fairness of the ML approach to several other targeting approaches (PMT, asset index, and geographic targeting), by evaluating each method's demographic parity, i.e., extent to which each method under- or over-targets specific demographic subgroups, relative to that group's true poverty rate (Figure 3c-d and Figure S9). Overall, we find that none of the targeting methods analyzed naively achieves parity across subgroups; a phenomenon referred to as "no fairness through unawareness."[28] We observe the most extreme parity differences in 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.e). 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 survey data indicate that 65% of adults and 85% of households have a phone; see also Figure S10); (*ii*) they must be a registered voter (approximately 81% of adults are registered to vote); (*iii*) they must self-target and attempt to register (approximately 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 (45% recall, per Table 1). Many of these sources of possible exclusion overlap; Table S6 thus estimates, based on the 2020 phone survey, the extent to which successive step in registration creates

additional exclusions. These results highlight the fact that algorithm-based targeting errors are a major 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 a social transfer 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 consideration are discussed in Methods §6. However, in the context of a crisis like the COVID-19 pandemic, it can be urgent to provide transfers quickly and with minimal in-person contact. Targeting via remote sensing and mobile phone data was more effective at reaching the poor than the feasible alternatives.

The results in this paper 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 cost-effective supplement that may be most useful in crisis settings or in contexts where traditional data sources are incomplete or out of date. Future work should explore how these complementary data sources can be best integrated in the design of inclusive social protections.

Methods

1. Related work

There is a rich history of theoretical and empirical work that compares and evaluates different 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 small 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] PMTs 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[37,38]. 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].

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

Togo is a small country of roughly 8 million in West Africa, where 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, increasing 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."[iii] Novissi was initially designed to provide benefits to informal workers in Greater Lomé, the large metropolitan area surrounding the capital city where

[iii] 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 3.6 million entries, which is roughly 81% of the total adult population, according to United Nations population estimates from 2019 (Table 2).

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 unique 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.e 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.[iv] 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 398 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 was very similar to the process 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.

[iv] 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 = 7,980).

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 were conducted in two wavs, with the first wave taking place between October and December 2018 and the second wave taking place between April and June 2019. We remove one observation that is missing consumption expenditure and asset data, for a total of 6,171 observations. Interviews took place with the head of household when possible, and if not with the most knowledgeable adult present, and lasted about three hours.

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 does not include an identifier for which member of the household the phone number belongs to. 4,201 households have phone numbers that place 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 = 4,201.

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 questions relating to 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. Note that since the sample is drawn based on inferred location, not all interviewees necessarily reside in an aid-eligible canton. The survey includes a question on canton of residence, and 76% 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 the dataset contains 7,980 observations corresponding to individual subscribers. We reweight the survey for nonresponse using the same CDR features and machine learning methods described in Methods §4. Specifically, we train a gradient boosting model (implemented with Microsoft's LightGBM) to predict response from CDR covariates, 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: Our main analysis relies on three measures of poverty: consumption expenditure (captured in the 2018-2019 field survey only), an asset-based wealth index, and a proxy-means test. The distribution of each outcome for the two survey datasets is shown in Figure S3.

- *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) accounting for sample weights. For the 2018-2019 field survey, the asset index is constructed from 24 underlying binary asset variables; for the 2020 phone survey it is constructed from 10 underlying binary asset variables. Note that 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 detailed in Table S1. The asset index explains 34.72% of the variance in asset ownership in the 2018-2019 field survey, and 49.06% of the variance in asset ownership in the 2020 phone survey.
- *Proxy-means test:* 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. 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[21,27]. 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.

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-specific poverty line of USD 1.79/day in each prefecture. Figure S4 shows the resulting poverty map. For validation, we evaluate the prefecture-level correlation between percentage poverty from the poverty mapping exercise and average consumption in the 2018-2019 field survey (though note that the latter survey is not representative at the prefecture level). 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 2km x 2km satellite tile in Togo.[12] We overlay the resulting tile-level wealth estimates with high resolution estimates of population density inferred from satellite imagery[39] to obtain population-weighted average wealth estimates for each canton, shown in Figure S4. 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 prefecture 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 7-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 antenna through which the cell tower 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 are 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 are a total of 4.89 million mobile network subscribers. We identify spammers on the network as any phone number that places 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: Between March 1 and September 30, 2020, we observe a total of 5.83 million mobile network 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[40] 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:* We calculate a set of features relating to mobile money usage. 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 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 4,201 match to CDR (91% 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 7,980 observations in the 2020 survey dataset are matched to CDR. Figure S3 compares distributions of selected CDR features grouped by ground-truth poverty metrics from the surveys.

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 surveys 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.[41,42] 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.[43,44] 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 past literature[13,14,20] to train models that predict poverty from CDR features. Specifically, we train a gradient boosting model 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 S3 and feature importances for each model are shown in Figure S6. The phone-based PMT explains approximately 41% of the total variation of a traditional PMT (Table S3, Panel C), as trained and evaluated using the 2020 phone survey dataset. When trained and evaluated using the nationally-representative 2018-2019 household survey with consumption data, we find that the phone-based PMT explains roughly 46% of the variation in total household consumption (Table S3, Panel A).

b. Parsimonious Method: "Total 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 "total 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.[v] We use these prices to back out the (approximate) amount spent by each subscriber in the time period from their outgoing mobile phone transaction logs. We find that the parsimonious method is significantly less accurate than the phone-based PMT, explaining 13% of the variation in both the 2020 phone survey PMT and the 2018-2019 household survey's consumption measure (Table S3, Panels A and C).

[v] 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. Targeting Methods

We simulate CDR-based and counterfactual targeting methods for reaching the poorest individuals in Togo, using our two survey datasets. For the 2018-2019 field 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 expenditure was not collected in the phone survey).

For each of the two datasets, we simulate four targeting methods that were available to the Togolese government in implementing the Novissi program: prefecture blanket targeting (targeting everyone in the poorest prefectures), canton blanket targeting (targeting everyone in the poorest cantons), occupation-based targeting, and CDR-based targeting. For prefecture blanket targeting and canton blanket targeting we targeting prefectures and cantons based on the corresponding poverty map, with the poorest areas targeted first. For occupation-based targeting, we categorize occupation of the respondent as recorded in each survey as either formal or informal/unemployed, using the same definition of informality used by the Novissi program.

We also simulate standard targeting methods that were not feasible for the Novissi program, as they require detailed data collection prior to program implementation. For the 2018-2019 field survey dataset we simulate targeting on the proxy-means test and asset-based wealth index; for the 2020 phone survey dataset we simulate targeting on the asset index only (since the PMT is used as ground truth). Note that 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 so that they can be fairly evaluated in targeting simulations.

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 use as a baseline should be considered an upper-bound.

b. 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.

- *Area under the ROC 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. 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 between 0% and 100%, we trace out the ROC curves shown in Figure S7. 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.

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 define 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.

c. 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 trace out 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 voter. We therefore assign each dataset a total budget of USD USD 25.83N, 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 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 2.

d. 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)[26] 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)[27] 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 four dimensions for parity: gender, ethnicity, religion, and age group. 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.

e. 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, 3.6 million unique adults possess a voter ID in Togo. Recent projections from the United Nations places the current adult population of Togo at 4.49 million[vi], for a voter ID penetration rate of 80%. In our 2020 phone survey dataset, which is concentrated in rural areas, 94.2% of respondents (weighted) report possessing a Voter ID.
- *Phone penetration:* In the 2018-2019 field survey, 65% of individuals report owning a mobile phone (Figure S10a); 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 S10b). These household estimates likely represent a lower bound, given the steady increase in

[vi] https://population.un.org/wpp/

phone penetration between 2018 and 2020. 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 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 S6.

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[45–49]; 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.e, many individuals in LMICs do not own mobile phones. Thus, any targeting system based on mobile phone data may exclude 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 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[50], and human rights doctrine emphasizes that any form of communications surveillance should be "necessary and proportionate"[51].

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 the IRB-approved researchers ever received access to CDR; data from the phone companies was never shared with either the Government of Togo or 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.

Data access: 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[46,52]. Yet such frameworks are still relatively nascent, and without careful consideration may exclude important stakeholders and perspectives[49].

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, we ex ante reweighted survey observations to ensure that the training data were representative of all mobile subscribers (Methods §3.a). We also perform ex-post audits to limit the likelihood that the trained models systematically disadvantage specific subgroups of the population (Methods §5.d).

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[53,54], and one that affects traditional proxy means tests and poverty scorecards[55,56]. However, recent evidence suggests that such distortionary effects may be limited[57], and complex eligibility criteria (such as the gradient boosting procedure described in Methods §4) should limit the scope for such gaming[58]. 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 nay be more appropriate[59].

Acknowledgements

Isabel Onate Falomir, Shikhar Mehra, Suraj Nair, Adrian Dar Serapio, Nathanael 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, Alex Nawar, and Daniel Quinn). We thank Esther Duflo, Luis Encinas, Tina George, and Ethan Ligon 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. Blumenstock is supported by NSF award IIS – 1942702. Authors retained full intellectual freedom in conducting this reserach, and as such all opinions and errors are ours

Data Availability

The data used to construct these estimates 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 is will be made publicly available shortly.

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. Kleinberg, J., Mullainathan, S. & Raghavan, M. Inherent Trade-Offs in the Fair Determination of Risk Scores. *arXiv:1609.05807 [cs, stat]* (2016).
27. 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.
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. Alderman, H. Do local officials know something we don't? Decentralization of targeted transfers in Albania. *Journal of Public Economics* **83**, 375–404 (2002).
38. Galasso, E. & Ravallion, M. Decentralized targeting of an antipoverty program. *Journal of Public Economics* **89**, 705–727 (2005).

39. Tiecke, T. G. *et al.* Mapping the world population one building at a time. *arXiv:1712.05839 [cs]* (2017).
40. 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).
41. 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).
42. 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.
43. 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).
44. 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.
45. 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).
46. Kerry, C. F., Kendall, J. & de Montjoye, Y.-A. Enabling Humanitarian Use of Mobile Phone Data. *Brookings Issues in Technology Innovation* (2014).
47. Blumenstock, J. E. Don't forget people in the use of big data for development. *Nature* **561**, 170–172 (2018).
48. Ohlenberg, T. *AI in Social Protection – Exploring Opportunities and Mitigating Risks*. (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2020).
49. 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.
50. Taylor, L. No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environ Plan D* **34**, 319–336 (2016).
51. Necessary and Proportionate. International Principles on the Application of Human Rights to Communications Surveillance. (2013).
52. 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.
53. Goodhart, C. *Monetary Relationships: A View from Threadneedle Street*. (University of Warwick, 1975).
54. 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).
55. Martinelli, C. & Parker, S. W. Deception and Misreporting in a Social Program. *Journal of the European Economic Association* **7**, 886–908 (2009).
56. Camacho, A. & Conover, E. Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy* **3**, 41–65 (2011).

57. 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.
58. Finkelstein, A. E-Z Tax: Tax Salience and Tax Rates. *The Quarterly Journal of Economics* **124**, 969–1010 (2009).
59. Björkegren, D., Blumenstock, J. E. & Knight, S. Manipulation-Proof Machine Learning. *arXiv:2004.03865 [cs, econ]* (2020).

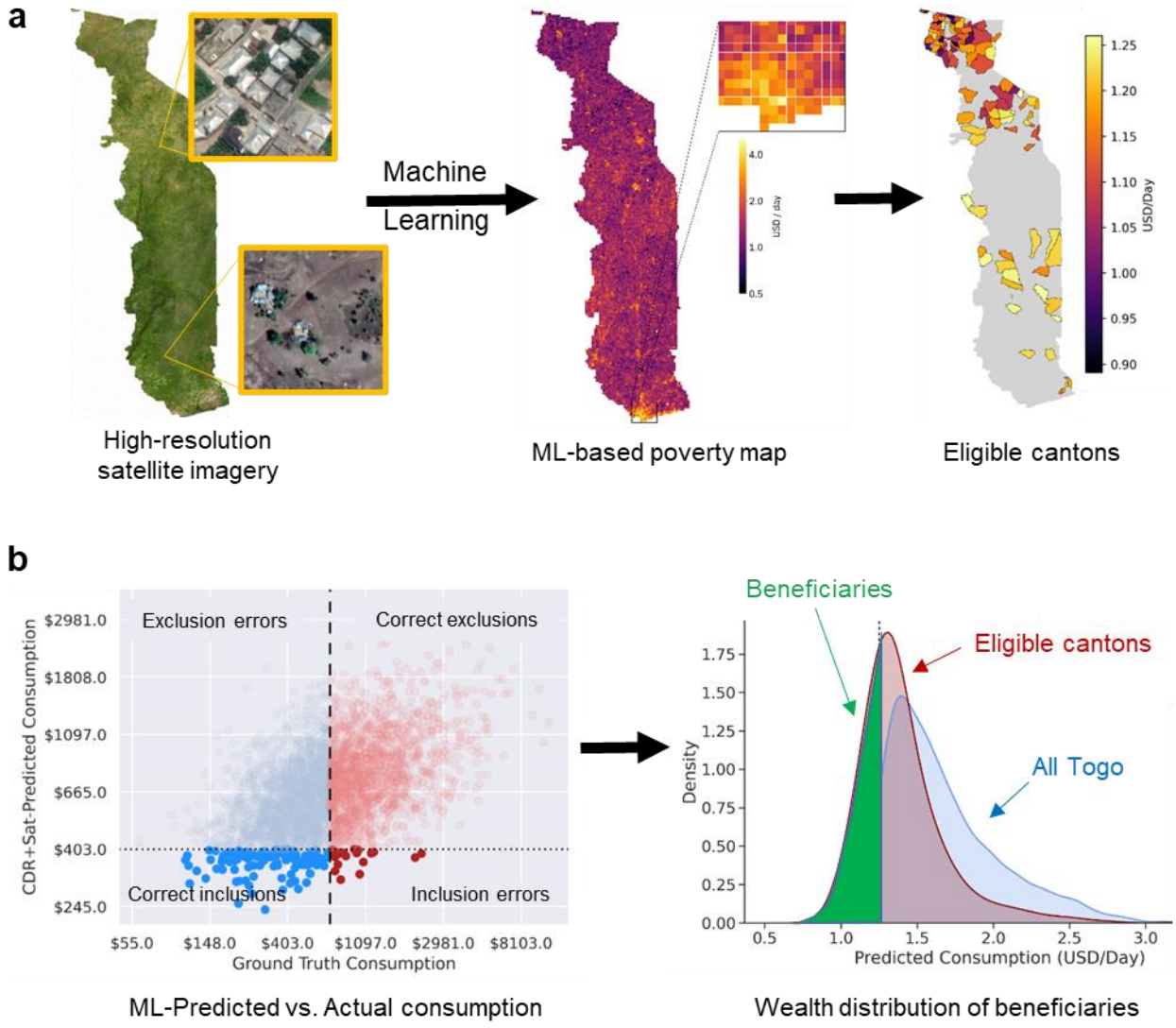


Figure 1 | Overview of methodology for rural Novissi. a) Selection of eligible cantons via satellite imagery: High-resolution satellite imagery (left) is combined with deep learning algorithms to generate poverty estimates for 2km x 2km grid cells (middle), which are overlayed with population data to produce canton level estimates of wealth. Voters registered in the 100 poorest cantons (right) are eligible for benefits. **b)** Selection of eligible beneficiaries via mobile phone data: A machine learning algorithm is trained based on representative survey data to predict consumption from features of phone use (Methods §4). The algorithm produces proxy poverty scores that are correlated with ground-truth measures of consumption (left). Subscribers who register for the program in targeted cantons with poverty scores over a threshold of USD $1.25/day are eligible for benefits (right). The blue distribution shows the predicted wealth distribution of the entire population of Togo; the red distribution shows the predicted wealth distribution in the 100 poorest cantons; and the green section indicates the predicted wealth distribution of Novissi beneficiaries. Estimates based on nationally representative survey data from 2018-2019.

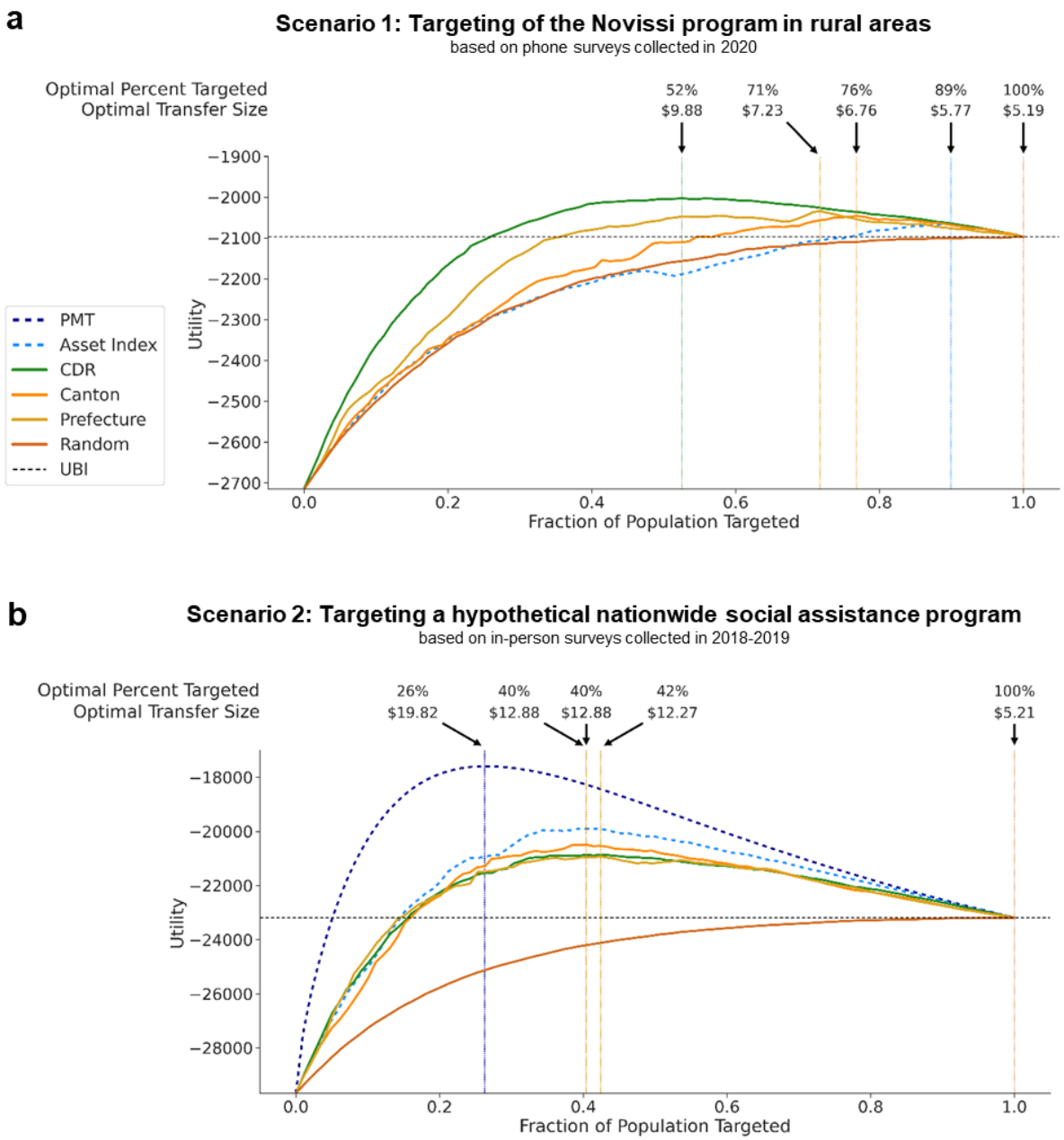


Figure 2 | Welfare analysis of different targeting mechanisms. Aggregate social welfare is calculated (assuming CRRA utility) under counterfactual targeting approaches. We assume a fixed budget of USD 4 million and a population of 154,238, with an equal transfer size for all beneficiaries. Utility curves for feasible targeting mechanisms are shown in solid lines; infeasible targeting mechanisms are shown in dashed lines. The horizontal dotted line indicates a universal basic income program that provides 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, using a 2020 phone survey of those cantons; and **(b)** a nationwide anti-poverty program, using a 2018-2019 nationally representative household survey.

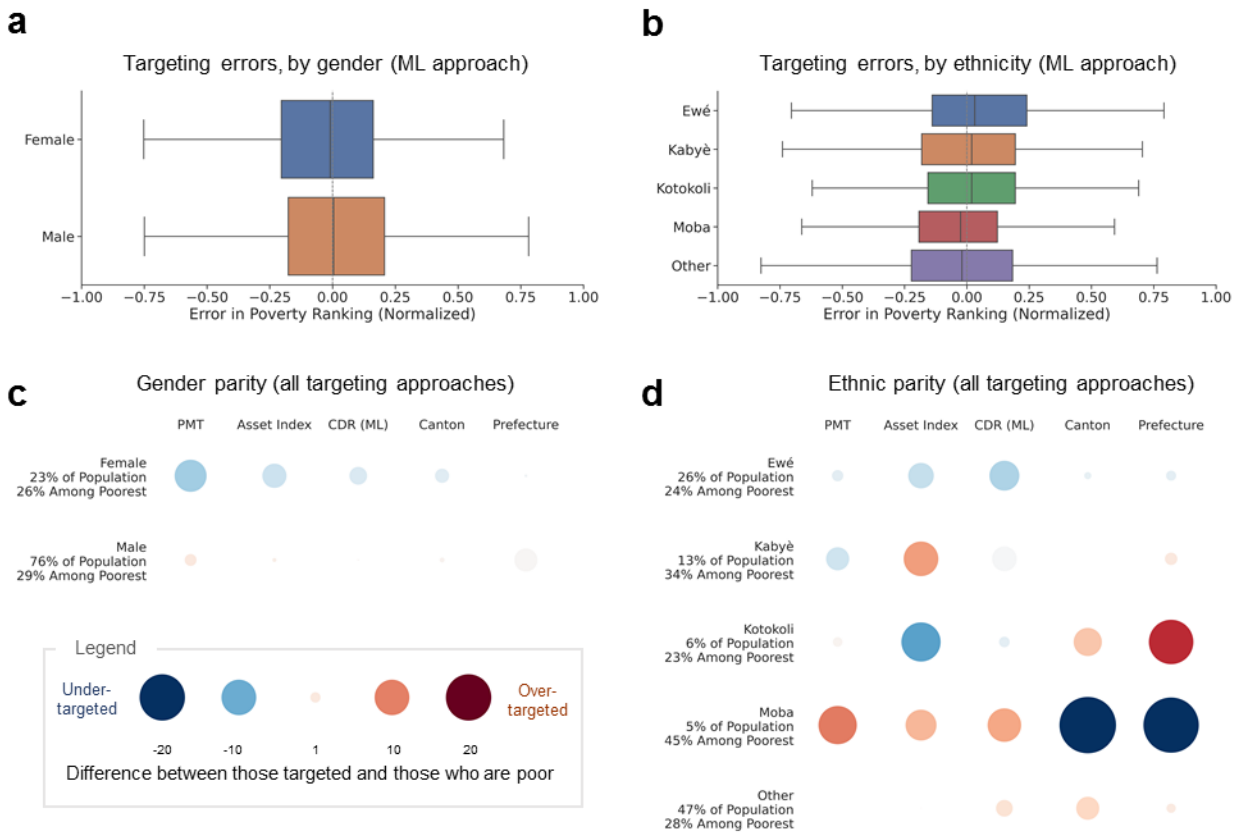


Figure 3 | Fairness of targeting for different demographic subgroups. We evaluate demographic parity across subgroups by comparing the proportion of a subgroup targeted under counterfactual approaches to the proportion of the subgroup that falls into the poorest 29% of the population (using data from the 2018-2019 field survey matched to CDR, N = 4,201). 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.

	2020 Phone Survey (N = 7,980)				*2018-2019 Field Survey (N = 4,201)*			
	Spearman	AUC	Accuracy	Precision & Recall	Spearman	AUC	Accuracy	Precision & Recall
Panel A: Targeting methods that were feasible in Togo in 2020								
Random	0.00	0.50	59%	28%	0.00	0.50	59%	29%
Occupation	--	--	55%	22%	--	--	58%	27%
Prefecture	0.24	0.57	67%	44%	0.27	0.64	68%	45%
Canton	0.29	0.58	66%	41%	0.31	0.67	69%	47%
Phone + ML	0.36	0.69	68%	45%	0.41	0.73	71%	51%
Panel B: Targeting methods that could not be implemented in Togo in 2020								
Asset Index	0.13	0.45	61%	33%	0.43	0.75	73%	53%
PMT	--	--	--	--	0.72	0.85	82%	68%

Table 1 | Performance of alternative targeting mechanisms. We simulate targeting the poorest 29% of the population according to the phone-based PMT and counterfactual targeting strategies, and evaluate how well each targeting method identifies the poorest 29% according to ground-truth poverty data. The first four columns evaluate targeting using a 2020 phone survey representative of subscribers in the 100 poorest cantons where Novissi was expanded; this analysis uses a Proxy Means Test to measure "true" poverty (since consumption data was not collected in the phone survey). The latter four columns evaluate targeting using nationally-representative household survey data collected in 2018-2019, where aggregate consumption is used to measure "true" poverty. Panel A compares the phone-based PMT (highlighted) to alternative targeting methods that could have been implemented in September 2020 when the Novissi program expanded; Panel B shows the performance of other targeting methods that were infeasible in Togo, since the government did not have asset or PMT data for the full population. We evaluate each targeting method using Spearman correlation coefficient, AUC, accuracy for reaching the poorest 29%, and precision and recall for reaching the poorest 29% (precision and recall are balanced in this application by construction). Table S4 and Table S5 present targeting results for alternative targeting thresholds, focusing on households below the poverty line and extreme poverty line rather than the poorest 29%.

Exclusion Source	Proportion	Data and Calculations
Voter ID possession	80.86%	According to government administrative data, 3,633,898 individuals are registered to vote in the country. The United Nations estimates the current adult population of Togo at 4,493,855 million, for a voter ID penetration rate of 80.93%.
SIM card and mobile phone access	65% - 85%	65% of individuals and 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.
Program awareness	40.40%	245,454 unique subscribers attempted to register for the rural Novissi aid program. However, we estimate that only 87% of these subscribers reside in an area eligible for the program. The total voting population of eligible areas is 528,562, for a 40.40% attempted registration rate.
Registration challenges	71.91%	Registration for the Novissi program requires completion of a short USSD survey. According to program administrative data, of the 245,454 subscribers who attempt registration, 176,517 (71.951%) eventually succeed, after an average of four attempts.
Targeting errors	55%	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 55%.

Table 2 | Sources of exclusion from rural Novissi benefits. We use program administrative data to estimate exclusion as a result of program requirements: voter ID possession (to provide identification), mobile phone ownership (to fill out the USSD platform), program awareness, and ability to register via a short USSD survey, as well as targeting errors. While this table assumes that these sources of exclusion are independent, Table S6 uses survey data to calculate overlaps in exclusions.

Supplementary Information

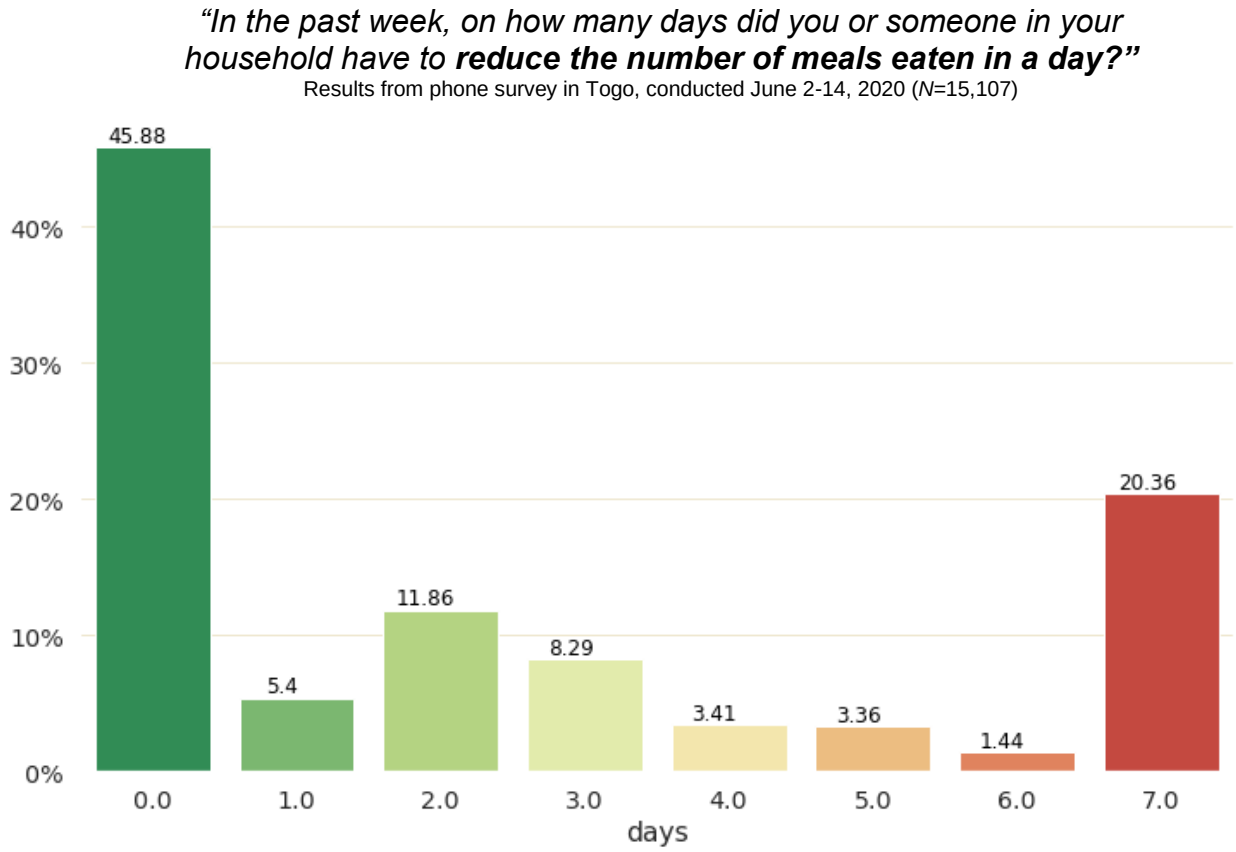


Figure S1 | Food insecurity in Togo. In June 2020, we conducted a phone survey of 15,107 mobile phone owners in Togo. Survey weights are used to make responses representative of the population of mobile phone owners in Togo.

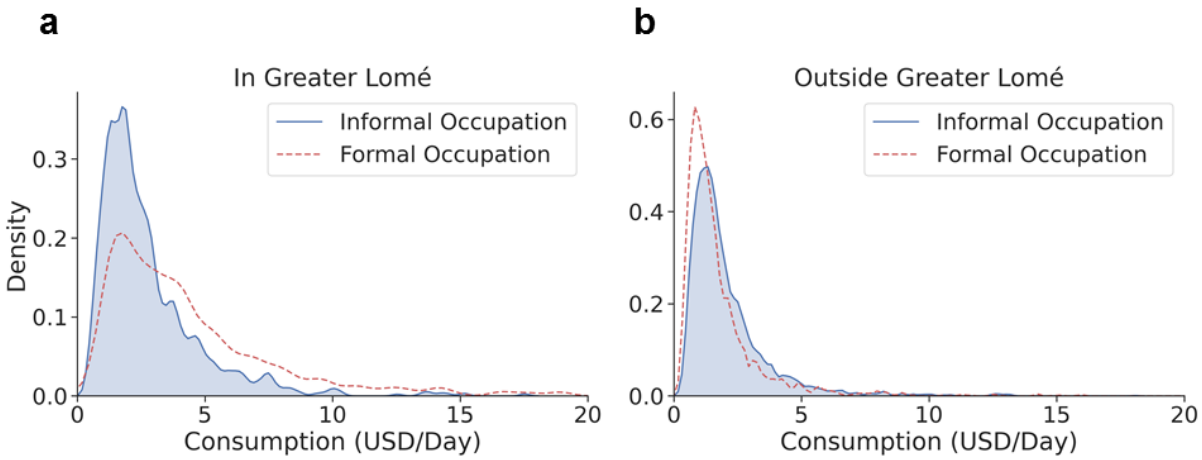


Figure S2 | Wealth of formal vs. informal workers. Results based on analysis of nationally-representative household survey data collected by the Government of Togo in 2018-2019 (*N* = 6,171). Data is collected at the household-level, we assign a household-level informal occupation indicator if at least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

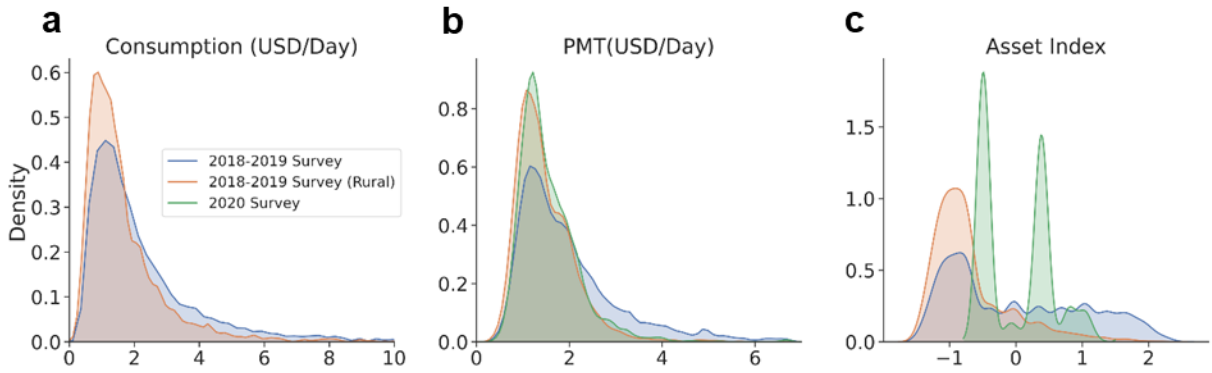


Figure S3 | Distributions of poverty outcomes. We show kernel density estimates for each of the poverty outcomes, separately for the 2018-2019 field survey (*N* = 4,201), for the 2018-2019 field survey restricted to rural households (*N* = 2,318), and for the 2020 phone survey (*N* = 7,980). Kernel density estimates are calculated with sample weights with a bin width of 0.1.

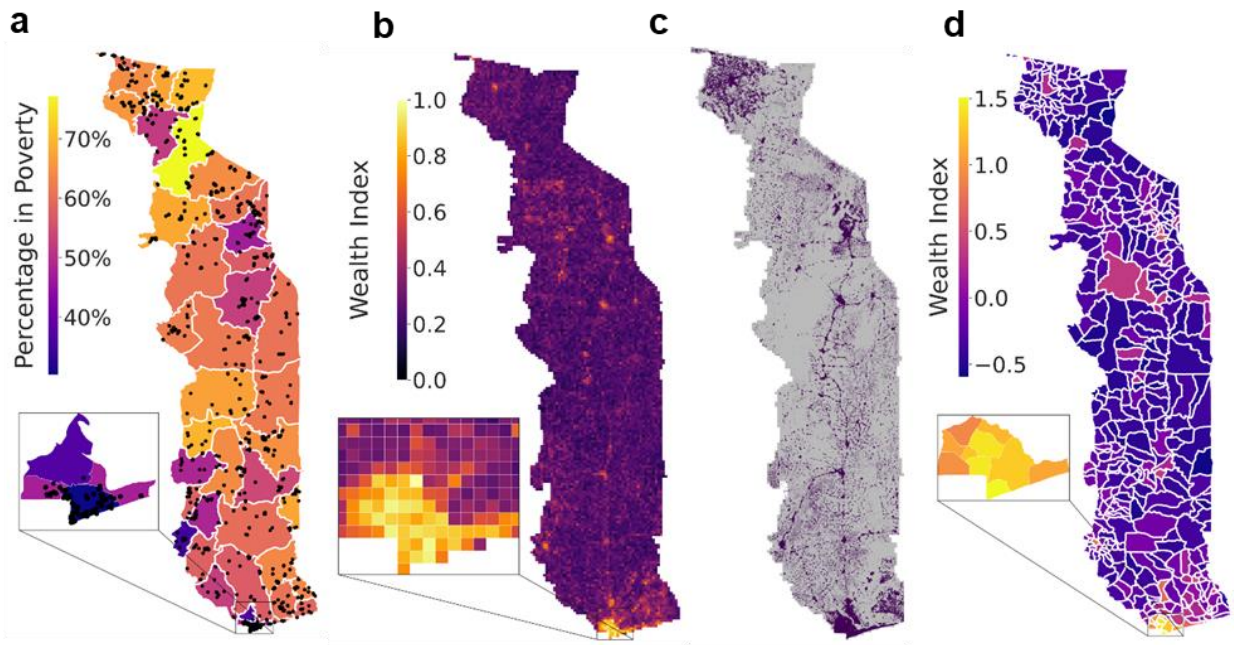


Figure S4 | Poverty maps. (a) Prefecture (admin-2) poverty map inferred from 2017 field survey (N = 26,902), showing the percent of the population living under the poverty line by prefecture. Overlayed with locations of survey observations in black points. (b) High-resolution estimates of consumption 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.

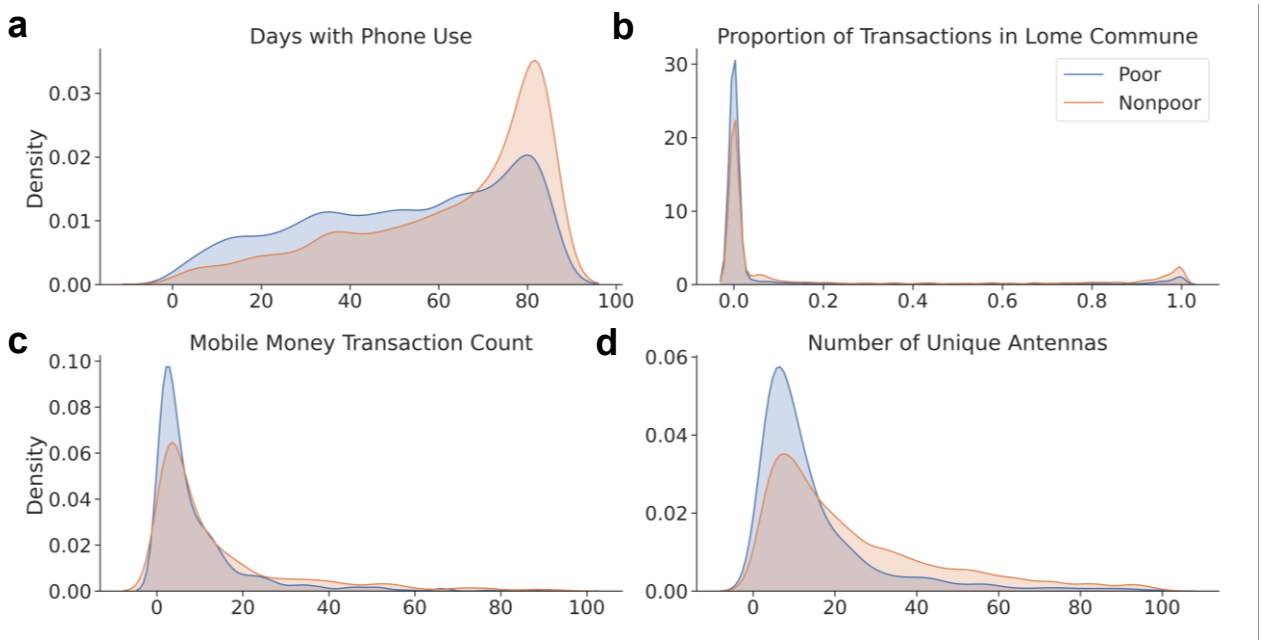


Figure S5 | CDR features. Comparing the distribution for CDR features for those above and below the international poverty line (USD 1.9/day) in the 2018-2019 field survey dataset.

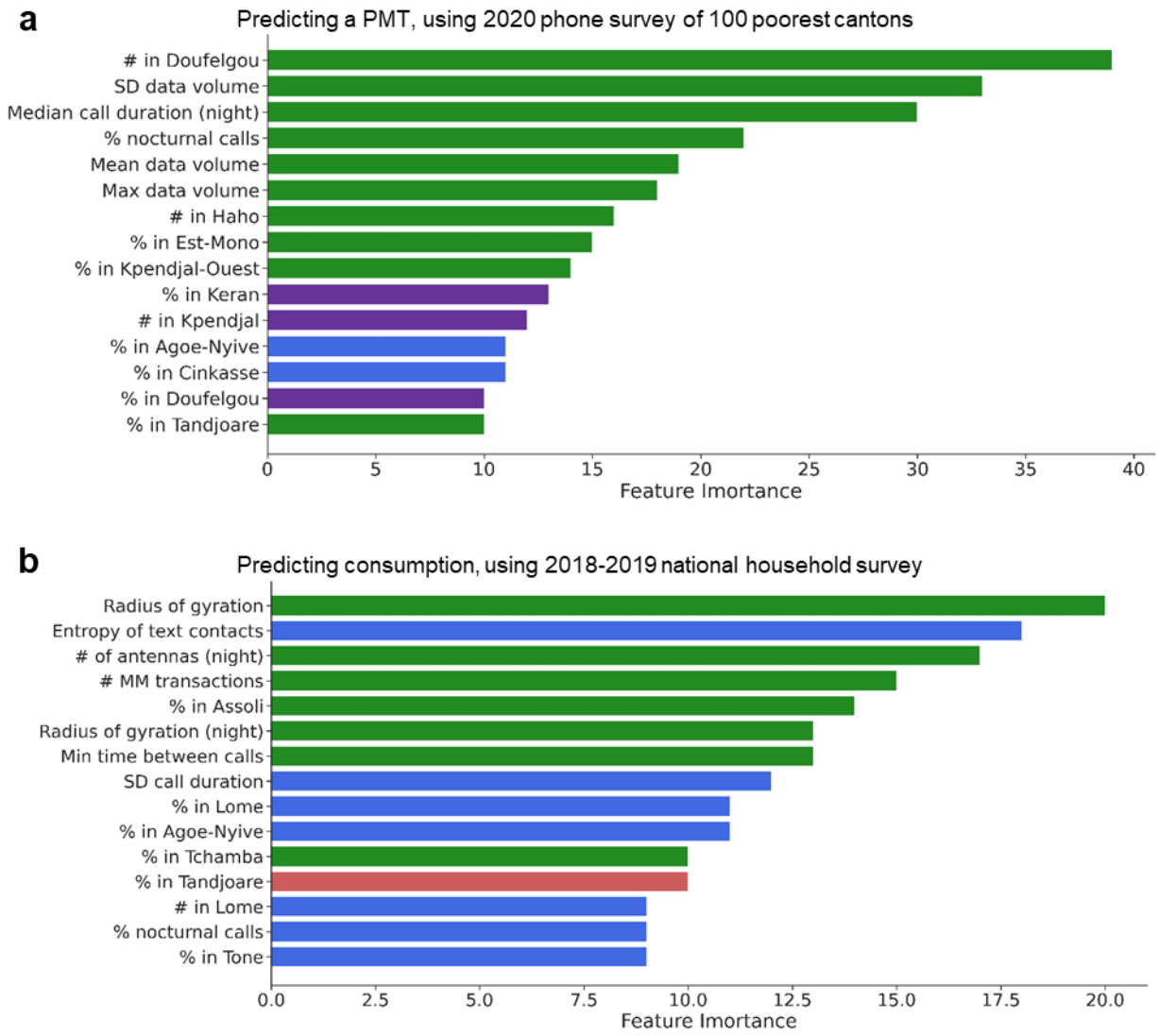


Figure S6 | Feature importances for ML models. Feature importances for the machine learning models trained to predict **(a)** Proxy Means Test from CDR, using a 2020 phone survey of mobile subscribers in Togo's 100 poorest cantons (N = 7,980); and **(b)** consumption from CDR in the 2018-2019 field survey dataset (N = 4,201). Feature importance is calculated based on the total number of times a feature is split upon in the random forest classifier. 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.

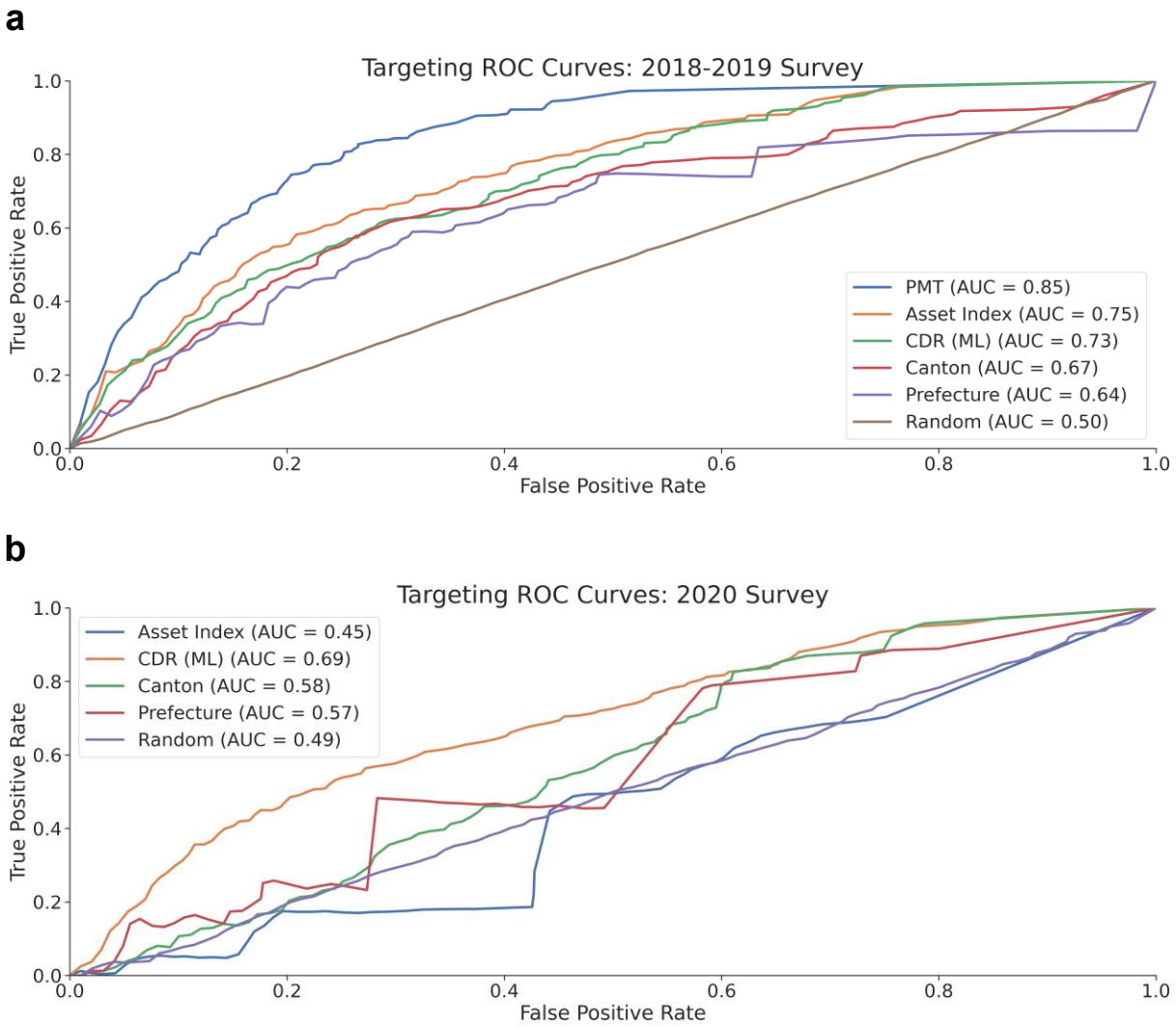


Figure S7 | ROC curves. ROC curves for counterfactual targeting methods using (a) 2018-2019 field survey dataset, and (b) 2020 phone survey dataset.

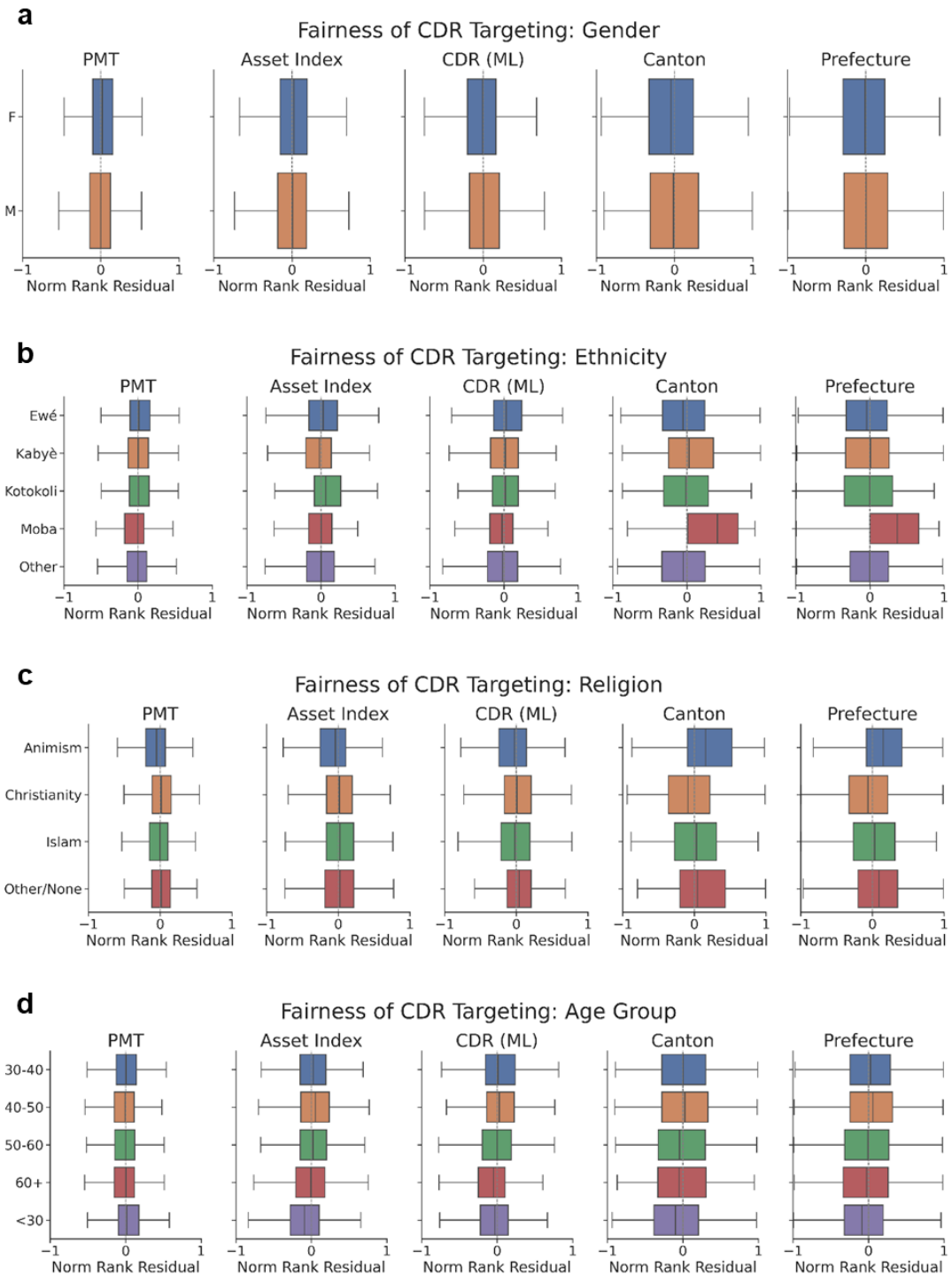


Figure S8 | Fairness with normalized rank residuals. Boxplots showing distributions of normalized rank residuals aggregated by subgroup. Left-shifted boxes indicate groups that are consistently under-ranked by a given targeting mechanism, right-shifted boxes indicate groups that are consistently over-ranked by a given targeting mechanism.

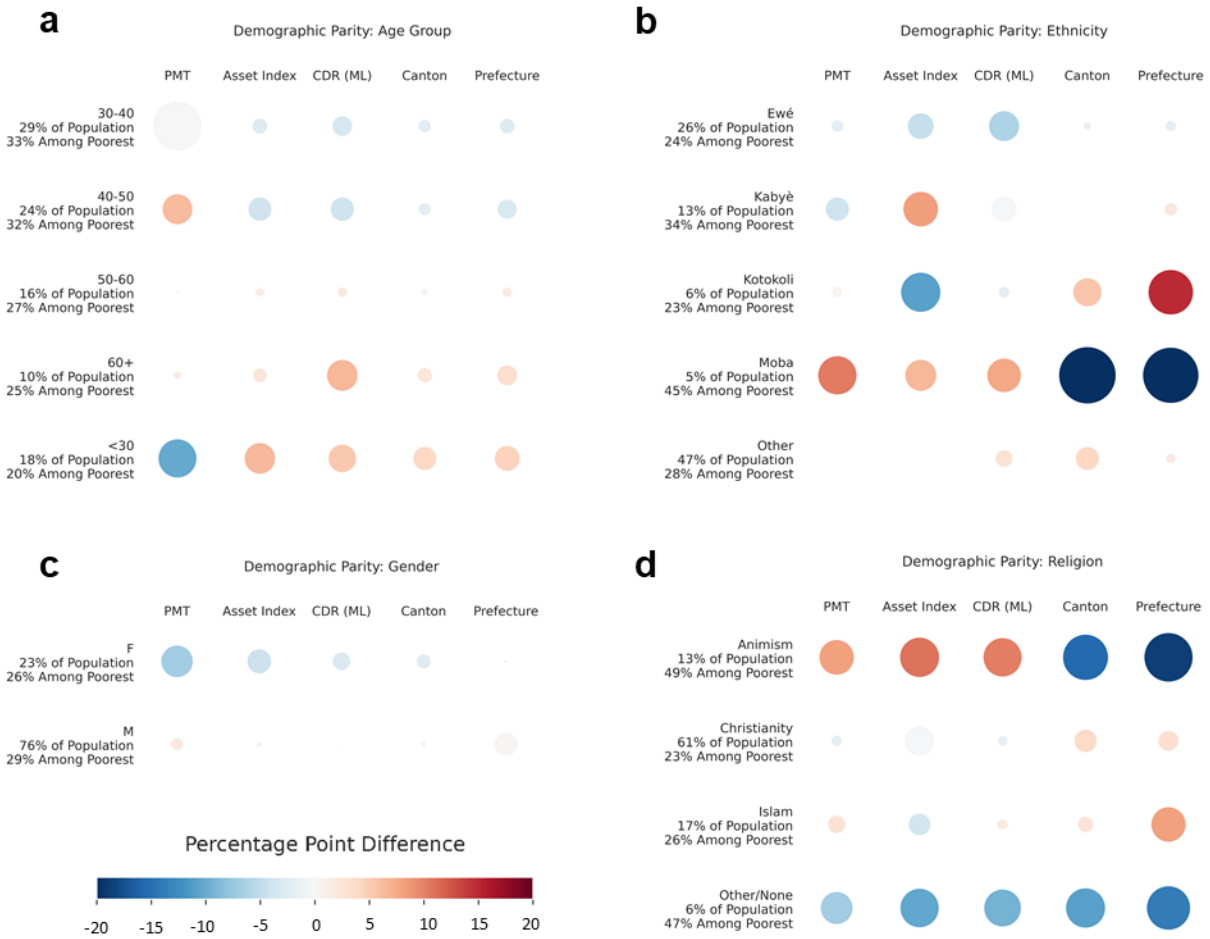


Figure S9 | Fairness of targeting mechamisms. We evaluate demographic parity across subgroups by comparing the proportion of a subgroup targeted under counterfactual approaches to the proportion of the subgroup that falls into the poorest 29% of the population (using data from the 2018-2019 field survey matched to CDR, N = 4,201). 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.

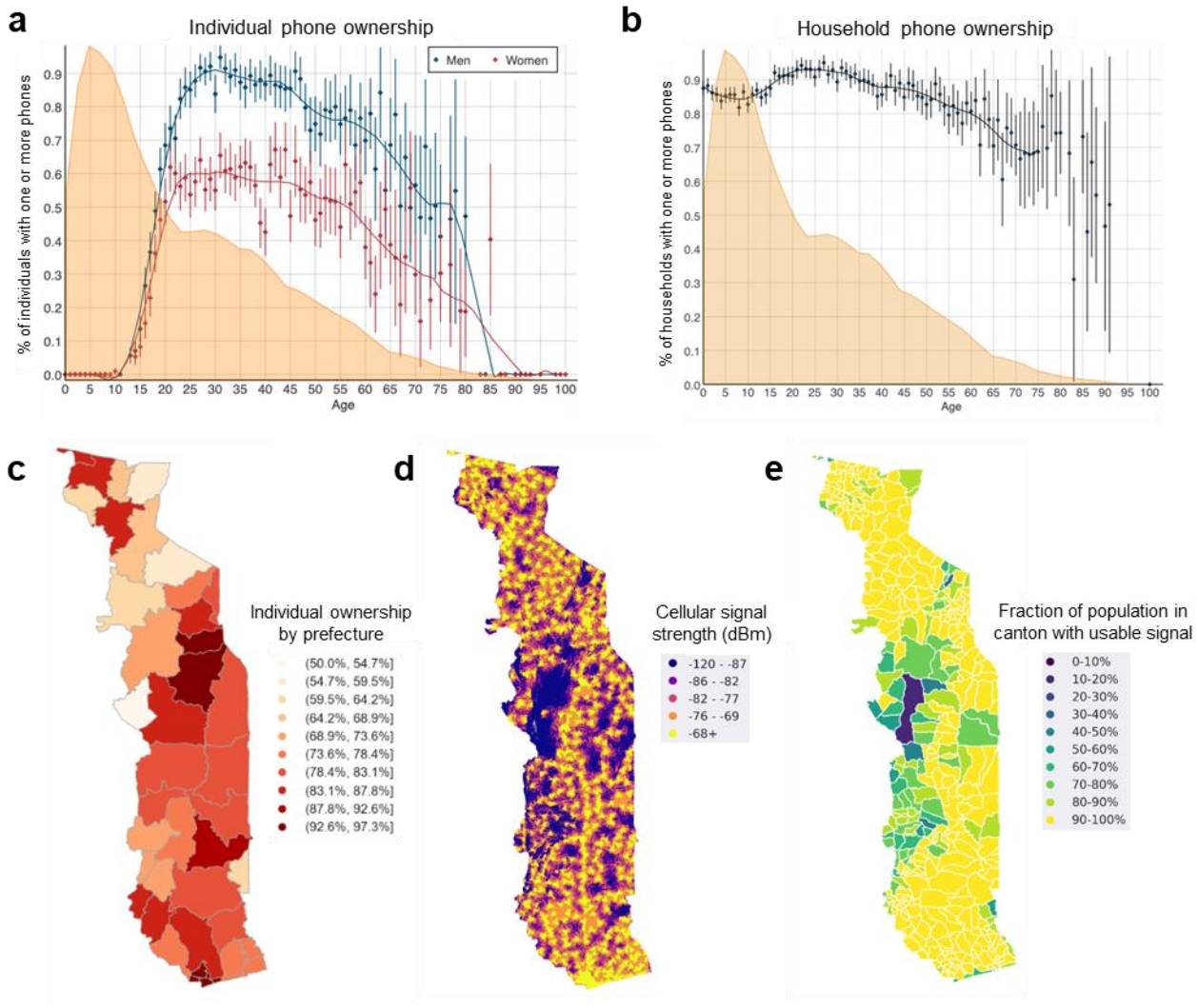


Figure S10 | Mobile phone penetration and coverage in Togo. Based on nationally-representative household survey data collected in 2018-2019, we estimate **a)** the percentage of adults in Togo with one or more mobile phone, disaggregated by age and gender (vertical bands indicate 95% confidence intervals); **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 downloaded from the Humanitarian Data Exchange.

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).

Asset	Consumption	PMT	Asset Index
Panel A: 2018-2019 Field Survey (N = 4,201)			
ML	0.46	0.62	0.74
Single Feature	0.13	0.16	0.11
Panel B: 2018-2019 Field Survey, Rural Only (N = 2,318)			
ML	0.31	0.44	0.51
Single Feature	0.09	0.12	0.08
Panel C: 2020 Phone Survey (N = 7,980)			
ML	--	0.41	0.40
Single Feature	--	0.13	0.14

Table S3 | Performance of phone-based PMT at 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.

	2020 Phone Survey (N = 7,980)			*2018-2019 Field Survey (N = 4,201)*		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Panel A: Targeting methods that were feasible in Togo in 2020						
Random	37%	81%	29%	50%	51%	29%
Occupation	34%	77%	27%	48%	48%	27%
Prefecture	42%	91%	33%	60%	69%	39%
Canton	41%	88%	32%	61%	70%	40%
Phone + ML	44%	93%	33%	63%	74%	42%
Panel B: Targeting methods that could not be implemented in Togo in 2020						
Asset Index	38%	82%	29%	65%	78%	44%
PMT	--	--	--	71%	89%	50%

Table S4 | Targeting households below the poverty line. We show the same targeting results as in Table 1, but evaluate to what extent targeting the poorest 29% reaches those below the international poverty line of USD 1.9 per person per day (76% of observations in the 2020 phone survey dataset and 57% of observations in the 2018-2019 field survey).

	2020 Phone Survey (N = 7,980)			*2018-2019 Field Survey (N = 4,201)*		
Targeting	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Panel A: Targeting methods that were feasible in Togo in 2020						
Random	48%	57%	29%	57%	35%	29%
Occupation	44%	50%	26%	56%	33%	28%
Prefecture	57%	74%	38%	66%	51%	43%
Canton	54%	68%	35%	67%	52%	44%
Phone + ML	58%	75%	39%	70%	57%	48%
Panel B: Targeting methods that could not be implemented in Togo in 2020						
Asset Index	49%	60%	31%	71%	60%	50%
PMT	--	--	--	80%	75%	63%

Table S5 | Targeting households below the extreme poverty line. We show the same targeting results as in Table 1, but evaluate to what extent targeting the poorest 29% reaches those below the 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).

Exclusion Source	N	Succeed	Drop Out	Percent of Total Remaining
All poor	2,004	--	--	100.00%
Own a voter ID	2,000	99.80%	0.20%	99.80%
Attempt to register for Novissi	1,548	71.59%	28.41%	71.44%
Succeed in registration	1,285	78.96%	21.04%	56.41%
Targeted by phone PMT	794	56.37%	43.63%	31.80%

Table S6 | Overlaps in sources of exclusion from rural Novissi. Attrition funnel for receiving benefits from the rural Novissi program, estimated with the 2020 phone survey dataset (N = 7,980). Percentages are calculated with sample weights.