

Poulin, Chloé et al.

Article

Performance of a novel machine learning-based proxy means test in comparison to other methods for targeting pro-poor water subsidies in Ghana

Development Engineering

Provided in Cooperation with:

Elsevier

Suggested Citation: Poulin, Chloé et al. (2022) : Performance of a novel machine learning-based proxy means test in comparison to other methods for targeting pro-poor water subsidies in Ghana, Development Engineering, ISSN 2352-7285, Elsevier, Amsterdam, Vol. 7, pp. 1-14, <https://doi.org/10.1016/j.deveng.2022.100098>

This Version is available at:

<https://hdl.handle.net/10419/299113>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Performance of a novel machine learning-based proxy means test in comparison to other methods for targeting pro-poor water subsidies in Ghana

Chloé Poulin^{*}, John Trimmer, Jessica Press-Williams, Bashiru Yachori, Ranjiv Khush, Rachel Peletz, Caroline Delaire

The Aquaya Institute, Nairobi, Kenya

ARTICLE INFO

Keywords:

Water subsidies
Poverty
Machine learning
Proxy means test
Community-based targeting
WASH

ABSTRACT

Access to safe drinking water is still very low among the poorest households in sub-Saharan Africa, and economic shocks can make water access even more difficult for poor consumers. Water subsidies can be a solution to enhance access to safe water services, but they are often ineffective as they regularly fail to reach the very poor. In this study, we developed a new Machine Learning-based proxy means test (ML-based PMT) to identify the poorest households and field-tested it in comparison to four other methods (the Demographic and Health Survey (DHS) wealth index, the Poverty Probability Index (PPI), Community Based Targeting (CBT) and the Ghana Government's Livelihood Empowerment Against Poverty (LEAP) program). We first developed our new ML-based PMT by applying machine learning techniques to the nationally-representative 2016–2017 Ghana Living Standards Survey and compared its performance with an existing PMT (the PPI). We then compared the strengths and weaknesses of this new method in three rural towns of southwestern Ghana against the four other methods, with respect to the characteristics of households they identified, their ease of implementation, their cost, and their acceptability among local stakeholders. In our field assessment we found that our new ML-based PMT performed better than most other approaches at screening out households having assets associated with wealth, but it had higher implementation costs than CBT and LEAP. Local government officials considered CBT to be more transparent than the PMTs, while community members perceived the PMTs to be fairer.

By highlighting the strengths and weaknesses of five different targeting methods, this study provides guidance to practitioners in choosing the most appropriate methods to target poor households eligible for water subsidies in rural Ghana.

1. Introduction

While access to improved drinking water supplies in sub-Saharan Africa has increased substantially over the past twenty years, water service providers and governments struggle to ensure that poor and vulnerable households benefit equally from these services. In Ghana, for example, the UNICEF/WHO Joint Monitoring Program (JMP) estimates that 94% of households in the richest quintile have access to basic water services, compared to only 54% among the poorest quintile, who often rely on surface water (25%), unimproved water supplies (9%), or water supplies located more than 30 min away (12%) (JMP, 2019). Poorer households face greater financial constraints related to water even when they do have access to piped supplies (Andres et al., 2019). Further, the

COVID-19 pandemic and the economic shocks associated with it have made water access more difficult for consumers in sub-Saharan Africa (WASHPaLS, 2020a). As a result, practitioners try to promote inclusive access to improved water sources and in particular to treated, piped water supplies.

Subsidizing access to safe water services for the poorest is critical to address this challenge. Although subsidies are common in the water sector, they are largely ineffective because they are often inadequately targeted, primarily benefitting high-income groups, and often failing to reach the very poor (Abramovsky et al., 2020; Andres et al., 2019). In addition, most of these existing subsidies apply to urban households with private piped connections and therefore do not reach the majority of poor households, who do not yet have piped water access. Notably,

^{*} Corresponding author.

E-mail address: poulin-chloe@hotmail.fr (C. Poulin).

<https://doi.org/10.1016/j.deveng.2022.100098>

Received 17 November 2021; Received in revised form 20 June 2022; Accepted 4 July 2022

Available online 5 August 2022

2352-7285/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

there are few examples of water subsidies in rural areas (Andres et al., 2019; Sasman, 2010), and there is no established approach to implement targeted water subsidies in these settings.

To target water subsidies to those that need them most, it is preferable to assess households' poverty status rather than simply targeting households based on their water source. First, there may be poor households who already use piped water (via private connections or communal standpipes) but either cannot afford sufficient quantities to meet their basic needs or spend a disproportionately large fraction of their income on it. Targeting based on non-piped water access would exclude them. Second, this targeting strategy could incentivize households who can afford to use piped water to wait for a subsidy. Finally, some rural water suppliers are promoting private piped connections in areas where communal standpipes are currently the only way to access piped water. In those cases, targeting subsidies based on non-access would be impractical as virtually no household has a private connection yet. As a result, various service providers and governments are proposing to target water and sanitation subsidies based on poverty status rather than access levels (Foster et al., 2000; Lauria et al., 2005; Hoo et al., 2022), including the Government of Ghana (Fosu, 2020; Ministry of Sanitation and Water Resources - Ghana, 2018). Existing methods to identify poor households include community-based targeting and survey-based proxy means testing.

Community-based targeting (CBT) is a participatory approach that involves asking community members to agree on a definition of poverty and to designate among themselves the households that meet this definition (Trémolet et al., 2010). Some forms of this method provide the community with a list of inclusion/exclusion criteria, while others ask the community to generate its own wealth rankings, sometimes using maps and cards to sort households into categories (Hargreaves et al., 2007; Trémolet et al., 2010). Generally, a perceived advantage of CBT is that community members have a more nuanced view of poverty levels than outsiders (e.g., they know which households do not receive any external support) and a better sense of long-term poverty, potentially increasing this method's reliability (Dershem et al., 2013).

A proxy-means test (PMT) involves predicting a household's poverty status based on a number of easily quantifiable proxy indicators of wealth. Example of proxy indicators include the dwelling's construction materials, education levels, types of assets owned, and types of food expenditures. The Demographic and Health Survey (DHS) wealth index is a common PMT derived by applying principal component analysis (PCA) to 10–15 survey answers (Rutstein and Johnson, 2004). The Poverty Probability Index (PPI), developed by Innovations for Poverty Action (IPA), is another PMT that relies on ten country-specific indicators identified by applying standard statistical learning techniques to national surveys. In Ghana, the PPI was derived using the nationally-representative 2016–2017 Ghana Living Standards Survey (GLSS7) (Kshirsagar et al., 2017). Importantly, those two PMTs did not use supervised machine learning methods, incorporated few variables, and may lack regional specificity.

In Ghana, the Ministry of Gender, Children and Social Protection used a combination of PMT and CBT to identify extremely poor households eligible for cash transfers and free health insurance under the existing Livelihood Empowerment Against Poverty (LEAP) program (Ashitey and Yeboah, 2012). This program involved conducting a PMT only among households selected via CBT and households that fell within specific vulnerability categories (elderly: >65 years old, disabled without productive capacity, orphaned, and vulnerable pregnant women and mothers with infants). The LEAP program has not been fully rolled out across the country yet: so far, 5079 communities out of approximately 72,790 communities (Ghana Statistical Services, personal communication, 2021) are covered by the program (Ministry of Gender, Children and Social Protection, 2022). Initial evaluations confirmed that LEAP beneficiaries were substantially poorer than the national average, with over 60% of them falling under the national poverty line (defined as the household's ability to afford a "consumption

basket" of 1760.8 GHS (approximately 292 USD) per adult equivalent per year, which allows meeting basic caloric needs) (Handa and Park, 2012; Akaligaung et al., 2016).

The relative performance of these approaches to identify households eligible for water subsidies in rural Ghana is currently unknown. Different studies in Indonesia (Alatas et al., 2016), Burkina Faso (Hillebrecht et al., 2020) and Cameroon (Stoeffler et al., 2016) found CBT to be somewhat less effective at identifying the poor than a PMT in rural areas, though the former led to higher community satisfaction (Alatas et al., 2012). However, other studies in nine different sub-Saharan countries showed that both CBT and proxy means testing had challenges targeting the most vulnerable, with a high risk of inclusion errors (i.e., truly non-poor households are incorrectly identified as poor) and exclusion errors (i.e., truly poor households are incorrectly identified as non-poor) (Aryeetey et al., 2010; Brown et al., 2018). Recent studies have used machine learning and satellite data to develop high resolution poverty maps (Chi et al., 2021; Jean et al., 2016; Yeh et al., 2020) using poverty indicators such as mobile phone use patterns or night-time lighting, but these maps do not have sufficient resolution for targeting water subsidies within a rural community. More broadly, geographic targeting may be useful in some urban settings, but likely not in small rural towns where spatial disparities in poverty and service levels are not as pronounced as in cities. Other studies have combined satellite data with mobile phone surveys and machine learning techniques (Aiken et al., 2020, 2021a, 2021b). However, these targeting techniques may omit households that do not have mobile phones, which are often the poorest households in low-income settings. For example, in Ghana, 18% of the households below the national poverty line do not own a mobile phone (GLSS7, 2018), increasing the risk of exclusion errors that would be associated with phone surveys.

In this study, we assessed whether a new Machine Learning (ML)-based PMT can perform better than CBT, existing PMTs (the DHS wealth index and the PPI), and the LEAP program (which is a hybrid approach) at identifying the poorest households in rural Ghana. We examined the strengths and weaknesses of the different approaches, offering insight into which methods should be used to implement pro-poor targeted subsidies (given stakeholder priorities). We hypothesized that our ML-based PMT would lead to fewer inclusion and exclusion errors than the conventional PPI and DHS wealth index, which are based on less sophisticated statistical techniques and do not take into account regional disparities within a country. We also hypothesized that the ML-based PMT may however be less acceptable to local communities and stakeholders than CBT, as it works as a "black-box" model and does not provide clear eligibility criteria.

The study focused on the Southern part of Ghana, where the social enterprise Safe Water Network (SWN) is planning to implement targeted subsidies as a way to address SDG 6.1 (achieving universal and equitable access to safe and affordable water for all by 2030). We conducted the field work for this study in three rural towns served by SWN. The findings from this study will inform SWN's program, and more broadly, will help promote more equitable access to safe water in rural Ghana.

2. Methods

Our work consisted of two main parts: (i) we first applied machine learning techniques to the GLSS7 dataset to derive a ML-based PMT predicting a household's official poverty status (as defined by the government based on the national poverty line), and we compared its performance on a subset of the GLSS7 dataset with an already existing PMT (the PPI); (ii) then we compared our new method with four other approaches in rural towns of southwestern Ghana. The field assessment (part ii) involved designing and deploying a survey-questionnaire that included questions from the ML-based PMT, the PPI, and the DHS wealth index, allowing us to estimate households' poverty status according to all three PMTs. It also involved collecting data on LEAP enrolment and implementing community consultations to identify poor households

according to CBT. We summarized the five different methods used in this study in Table 1.

2.1. Computational development and validation of the ML-based proxy-means test

2.1.1. GLSS7 dataset cleaning

We used the GLSS7 dataset (GLSS7, 2018) to derive our ML-based PMT. This nationally-representative survey on economic conditions

Table 1
Summary of the five different poverty targeting methods examined in this study.

| Method | Definition of poverty | Part-1 comparison | Part-2 comparison |
|------------------|--|--|--|
| Reference | The household is below the national poverty line defined by the Ghana Statistical Service (consumption below 1760.8 GHS per adult equivalent per year) as measured by the GLSS7. | | |
| ML-based PMT | The household's poverty probability, as derived using machine learning from 47 indicators, is within the highest quintile. ^a | Desk-based: Compared poverty predictions against the reference on a subset of households from the GLSS7 dataset (2539 households in the test set). | Field-based: Compared the characteristics of households identified as poor within a study population of 818 households; also assessed costs, ease of implementation, and acceptance by local stakeholders. |
| PPI | The household's poverty probability, as derived using IPA's existing method (linear combination of 10 country-specific indicators), is within the highest quintile. ^a | | |
| DHS wealth index | The household's wealth index, as derived using the DHS method (linear combination of 142 indicators derived from 13 survey questions) is within the highest quintile. | | |
| CBT | The household was designated as vulnerable ^b during a community consultation. | | |
| LEAP | The household reported being a beneficiary of the national poverty alleviation LEAP program (which used a combination of CBT and PMT with three other vulnerability criteria). | | |

^a For each PMT, we set a cut-off threshold to include 20% of the households with the highest probability of being poor. Using the same cut-off allowed for direct comparisons between the three PMTs, and 20% was a reasonable percentage for targeted subsidies.

^b According to our field protocol, a household was designated as vulnerable by the community if (i) members of the household could not feed all year round without help from relatives, or (ii) the household included an elderly person (>65 years old), a person having a severe disability or a chronic illness preventing them from working, a widow, or an orphan, and the household did not receive support from relatives.

was administered in 2016–2017 by the Ghana Statistical Service. A total of 14,009 household interviews were conducted across the country, collecting data on hundreds of variables covering general topics such as demographic characteristics, education, health, employment, housing, income, and expenditures.

There are vast disparities in poverty between the Southern and Northern areas of Ghana (15% of households fall below the national poverty line in the South, versus 56% in the North). As this study focused on Southern Ghana, we removed the Northern, Upper East, and Upper West regions from the GLSS7 dataset (which fall within the “Northern” category). We also removed the Greater Accra Region as it is a megapolopolis and not comparable to more rural areas or small towns. We kept both rural and urban households in the GLSS7 dataset; Ghana defines as urban any settlement with more than 5000 people, which therefore includes a large portion of small towns serviced by our implementation partner, Safe Water Network. We then extracted variables related to household characteristics, asset ownership, expenditures, and water access from the entire survey as possible candidate indicators to include in our ML-based PMT. We transformed each categorical variable into multiple binary (yes/no) indicators. The final reduced dataset for the Southern part of Ghana included 8464 households and 623 indicators total.

The GLSS7 dataset also contained a binary variable indicating households' official poverty status as defined according to the national poverty line (consumption below 1760.8 GHS or 287 USD per year per adult equivalent). We used this binary poverty variable as our reference against which we trained our ML-based PMT. Note that we used the version of the PPI derived against the same reference poverty status, allowing a direct comparison of performance with our ML-based PMT.

2.1.2. Feature reduction

The aim of feature reduction is to select a subset of relevant indicators that will allow good model predictions by reducing overfitting and training time. Due to the massive growth of data, feature reduction techniques have been used across many scientific disciplines (Cheng et al., 2021; Karasu et al., 2020; Remeseiro and Bolon-Canedo, 2019). In this study, we used feature reduction techniques to select the most relevant subset for predicting poverty from the 623 indicators in our dataset, and we used R software (R version 6.3.6) to conduct the analysis.

Before applying machine learning algorithms for feature reduction and to ensure that included variables could be easily measured in a survey, we first manually removed collinear indicators from the regional dataset by using the Variance Inflation Factor (VIF). We kept indicators where $VIF < 5$ (Chatterjee and Simonoff, 2013; O'Brien, 2007); then we verified with a correlation matrix that all remaining pairs of indicators had linear correlations in which $R^2 < 0.7$ (Taylor, 1990). When R^2 was above 0.7 we included the indicator that was simpler to ask in a questionnaire (for example, between “do you trust members from other ethnic groups” and “do you trust people in your village”, we chose the latter to keep in the survey). This first reduction method decreased the number of indicators from 623 to 491 (Fig. 1). We then used the LASSO (Least Absolute Shrinkage and Selection Operator) algorithm (Tibshirani, 1996) as part of machine learning to further reduce the number of indicators (for more details see Appendix A.1). With the LASSO feature reduction method, the number of indicators was reduced from 491 to 47 (i.e., 47 indicators had non-zero β coefficients) (Fig. 1).

2.1.3. Predictive model

Predictive modelling is used to estimate the probability of an outcome based on a set of predictor indicators. This practice is widely used in diverse domains of science (Ahmad et al., 2018; Podgorski and Berg, 2020; Rao et al., 2017; Rodriguez-Galiano et al., 2018; Thoe et al., 2014) and to predict poverty at different resolutions (Perez et al., 2019; Zhao et al., 2019). In this study, we developed two models with two different sets of predictors to estimate poverty at the household level. In

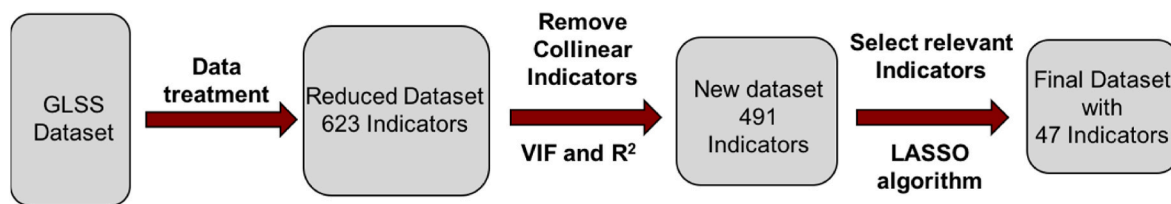


Fig. 1. Different feature reduction steps applied on the GLSS7 dataset to obtain a final reduced set of 47 indicators that we included in our ML-based PMT.

the first model, we took all 47 predictors selected by the LASSO algorithm ($\beta \neq 0$). For the second model, we took the first ten predictors among the 47 pre-selected predictors. The LASSO algorithm ranked the variables according to their β coefficients: we took the ten variables whose β coefficients had the highest absolute values, meaning that they had the strongest relationships with the poverty variable. We used the model with ten predictors as a point of direct comparison with the PPI, which also used ten predictors.

We split the 8464 households of the regional dataset into a training set (70% of households, to develop the model) and a test set (30% of households, to validate the model), and we ensured that each set contained the same proportion of poor households as the original proportion in the full Southern GLSS7 dataset (i.e., 15% of the households).

Using the training set, we applied the SuperLearner (Laan et al., 2007) metalearning algorithm in R to develop our predictive model. This metalearning algorithm used cross validation to estimate the performance of multiple machine learning models, or the same model with different settings. It then created an optimal weighted average of those models called an “ensemble”. We built an ensemble by using six different machine learning algorithms (Appendix A – Table A.1). We used a 10-fold cross validation and the Area under the Receiver Operating Characteristic curve (AUC) as a meta-learning method to combine the different algorithms because we had an unbalanced binary outcome (LeDell et al., 2016) (i.e., the proportion of poor was substantially lower than the proportion of non-poor). We then trained the two different models (with 47 and 10 predictors). Each model computed a household’s probability of poverty. To convert this output probability to a binary (poor/non-poor) outcome, we set a cut-off threshold such that the 20% of households with the highest probability were considered to be poor. The remaining 80% were considered to be non-poor (Table 1). This 20% cut-off threshold i) allowed for a direct comparison between the ML-based PMT, the PPI, and the DHS, and ii) was a realistic quota for program implementation, as threshold selection should try to replicate actual program coverage as much as possible (Schnitzer and Stoeffler, 2022).

In addition to the metalearning algorithm, we also tested a random forest model and a multi-linear regression model for comparison (Appendix A.3).

We then assessed the performance of each model on the test set (using the GLSS7 national poverty status as the reference), by calculating its accuracy, sensitivity, specificity, and AUC (all terms are defined in Appendix A.2). Using similar methods, we also trained and tested a separate model on the national GLSS7 dataset (12,611 households, still excluding Greater Accra Region) to allow for a more direct comparison with the PPI, which was also developed at the national scale. The method and the results of this national scale model are presented in Appendix A.4, while the main text primarily focuses on the regional models.

2.1.4. PPI score

We computed PPI scores for households in the test set to compare the PPI’s performance with our ML model. The PPI involves asking 10 questions about household’s characteristics and asset ownership, and then deriving a score to compute the likelihood that the household is living below the poverty line (Kshirsagar et al., 2017). We extracted the

10 indicators needed for the PPI score from the reduced GLSS7 dataset and used the 2016 PPI scorecard corresponding to Ghana’s National Poverty Line (our reference poverty definition) to derive the PPI score for each household in the test set (PPI, 2019). Each PPI score was then converted into its corresponding poverty likelihood using a look-up table (a lower PPI score indicates a higher probability for the household to be poor) (PPI, 2019). To convert the poverty probability into a binary variable (i.e., poor and non-poor), we used the same procedure as for the ML-based PMT to set the probability cut-off. A household was considered to be poor if it fell within the quintile with the highest poverty probability. We then assessed the performance of the PPI prediction by calculating its accuracy, sensitivity, specificity, and AUC.

2.2. Field-based assessment

2.2.1. Field data collection

Between March and May 2021, we conducted field data collection in three rural towns of the Ahafo and Ashanti regions: Gambia No. 2 (Asutifi North district), Anyinasuso (Ahafo Ano North municipality), and Breme (Tano South municipality). These three towns are service areas of Safe Water Network (SWN), a social enterprise providing water services across rural Ghana and partner of this study. We selected these three towns among SWN’s 92 service areas because i) they were located in the funder’s (Conrad N. Hilton Foundation’s) priority regions, and ii) SWN shortlisted them for possible future implementation of targeted water subsidies. Through community consultation meetings and household surveys (Fig. 2), we collected information for the following five methods: the regional ML-based PMT, the PPI, the DHS wealth index, LEAP and CBT. We also collected information on stakeholder and community members’ perceptions regarding the different methods.

Two of the three towns had populations of approximately 3000–5000 people, and the third had approximately 1500 people, though exact population data were not available. Across the three towns, we delineated neighborhoods of less than 300 households, the preferred maximum size for conducting community consultation meetings according to our prior experience (WASHPaLS, 2020b). Because the third town contained approximately 300 households, we surveyed all households. Overall, we selected a total of six neighborhoods covering 877 households.

For the CBT method, our team coordinated a community consultation meeting in each neighborhood. This required training District Assembly staff on our established protocol (WASHPaLS, 2020b) and liaising with local leaders to make repeated announcements to neighborhood residents. On the meeting day, our team was present to support District Assembly staff as needed, help ensure that meeting participants were exclusively from the target neighborhood, and monitor fidelity to the protocol (see the detailed protocol in Appendix B). Community consultation meetings had 40–120 participants (with equal representation of genders) and lasted approximately one hour. To determine whether a household was poor or not, the field team was trained to structure communities’ answers into the two following categories: (i) households that could not feed all year round without help from relatives, or (ii) households that included an elderly person (>65 years old), a person having a severe disability or a chronic illness preventing them from working, a widow, or an orphan and did not receive support from

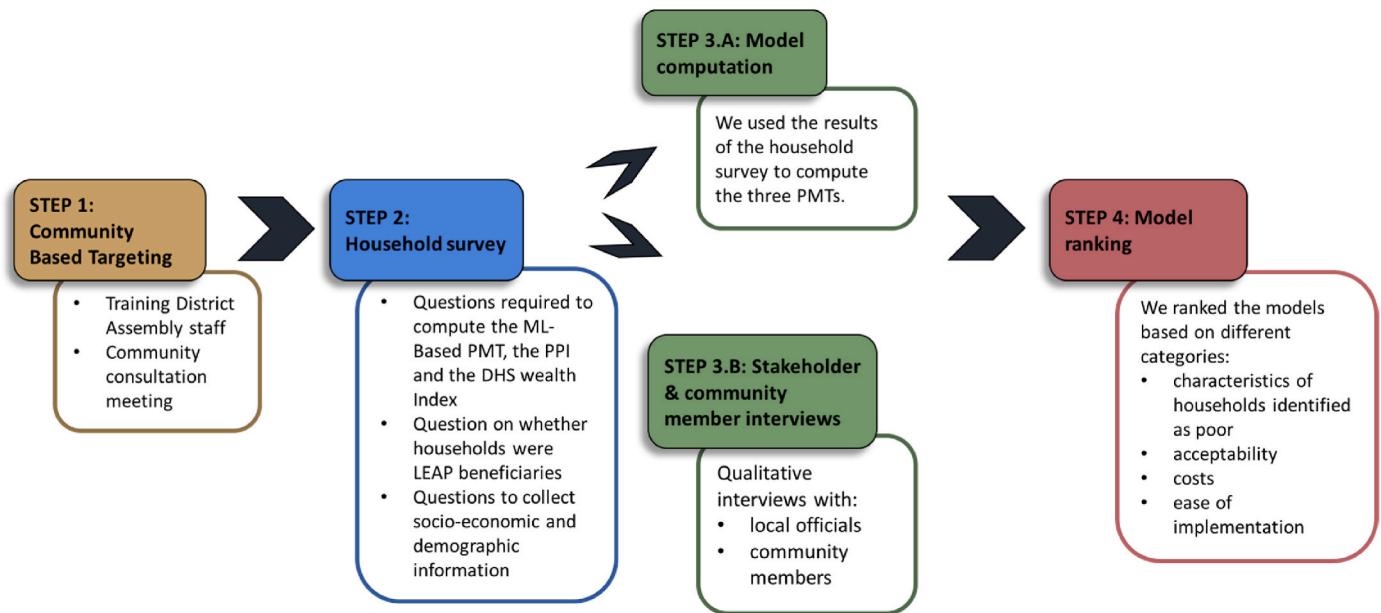


Fig. 2. Steps involved in the field-based assessment of poverty identification methods.

relatives. These categories correspond to characteristics that communities in Ghana typically cite when asked about poverty (WASHPaLS, 2021, 2020b) and also to characteristics cited in Ghana's national pro-poor guidelines (Ministry of Sanitation and Water Resources - Ghana, 2018).

Following each community consultation meeting, three enumerators surveyed all eligible households in the neighborhood (households were eligible if an adult at least 18 years of age was present once over the course of three visits). The survey included questions capturing the 47 household characteristics required for our regional ML-based model, as well as additional questions required to compute the PPI and DHS wealth index. We also collected demographic and socio-economic information, as well as whether households were LEAP beneficiaries (see the final survey in Appendix C). Enumerators obtained written informed consent from all survey participants, administered the survey in the local language (Twi), and recorded responses in the CommCare mobile phone application (version 2.45, DiMagi Inc). For quality control, we conducted spot checks or back checks on 5% of surveys. We also reviewed all answers to a subset of survey questions daily and followed up with enumerators to clarify inconsistent responses.

2.2.2. DHS wealth index computation

The DHS wealth index is based on a set of indicators that include household assets, dwelling characteristics, and utility services (Rutstein and Johnson, 2004). A Principal Component Analysis (PCA) is used to assign a weight to each standardized indicator, and then the sum of the weighted indicators produces the household's wealth index value. The distribution of the household population (where each member is given his/her households' index score) is divided into quintiles and a poverty score is attributed to each quintile (Rutstein and Johnson, 2004). We replaced any missing values with the indicator's average in our dataset, and we applied the already computed DHS-Ghana scores for rural areas to derive each household's wealth index (Rutstein, 2008). Finally, we used the 20th percentile as a poverty threshold to turn the results into a binary score (i.e., households below the 20th percentile were considered to be poor).

2.2.3. Methods comparison

We compared the five methods with regard to the characteristics of households they identified as poor (whether they had characteristics suggestive of greater vulnerability and less wealth, and whether they

faced greater issues with water access), their scalability (ease of implementation and affordability), and their acceptability to local stakeholders (community members, district and national authorities).

2.2.3.1. Outcomes of household identification. The collection of the large amount of complex information needed to compute total household expenditures and determine whether a household was poor according to the national poverty line (GLSS7, 2018) was beyond the scope of this work. We instead used vulnerability and wealth checks that practitioners and local stakeholders might instinctively apply in the field to quickly assess whether the methods are selecting households that align with their expectations. In other words, we used these checks to determine whether local stakeholders would likely perceive targeted households to be more vulnerable than others.

We used the GLSS7 dataset to determine the indicators to be employed as wealth checks. We examined Pearson correlation coefficients between 64 indicators (regarding housing and assets) and annual expenditures per adult equivalent. We identified nine indicators having an absolute correlation coefficient above 0.2 (see Appendix D): ownership of a refrigerator, a stove gas, an electric iron, a fan, stuffed furniture, a car, a television, a laptop computer, and a bank account or loan scheme. To determine indicators to be employed as vulnerability checks, we turned to the Government of Ghana's guidelines for targeting poor households for basic sanitation (Ministry of Sanitation and Water Resources - Ghana, 2018) and to prior research studies in rural Ghana (WASHPaLS, 2021). We retained three indicators: household head is female; household head is elderly, has a disability, or has a chronic illness; and household cannot feed all year round without the support of neighbors or relatives. The first two indicators came from the Government of Ghana's guidelines for targeting poor households (Ministry of Sanitation and Water Resources - Ghana, 2018). The third indicator came from our extensive prior work in rural Ghana on identifying the poor for sanitation subsidies, which involved over a hundred consultations with communities and local government officials to establish a locally appropriate definition of poverty (WASHPaLS, 2020b). These vulnerability and wealth checks were meant to represent "intuitive indicators" that field practitioners would likely use as verification checks to quickly assess whether targeting methods made sense. They were not meant to definitively determine whether one method is better than another: the wealth checks are based on assets and will favour models such as the DHS that were based on asset ownership, while the PPI and

the ML-Based PMT were based on consumption. In other words, we used these checks to determine whether local stakeholders would likely perceive that the households identified as poor by the methods were more vulnerable than the rest, and that households identified as non-poor were wealthier. These checks were thus a way to assess acceptability among local stakeholders and communities.

In addition to vulnerability and wealth checks, we used “water checks” to determine the extent to which households identified as poor were also those most in need of safer water access. With these “water checks” we wanted to confirm whether targeting water subsidies based on poverty only would be effective or if including additional water criteria may be helpful. Our survey included questions about water access, water consumption, and challenges that households typically face in accessing water (e.g., poor water quality, waiting in queues, unreliability of the water supply, or financial constraints to get water). We selected four indicators reflecting different aspects of drinking water access: source of drinking water, time to fetch drinking water, water consumption per person per day, and self-reported challenges in accessing safe water. We then compared these four variables among households identified as poor and non-poor by each method. It is important to note that the three study sites may not be representative of water access in rural Ghana generally, especially because Safe Water Network’s presence might have increased access to safe water.

Finally, we used the above vulnerability checks to determine the characteristics of households left behind by the methods (i.e. identified as non-poor) but still reporting a financial constraint to get water.

2.2.3.2. Scalability and acceptability. We documented the cost of implementing each poverty identification method, including personnel costs (stipends and per-diems), transport, and logistics (communication, printing, chair rental for community meetings). For the PMT methods (which all fell under survey-based methods), we estimated that some of these costs would be lower with a shorter survey (“variable costs” like personnel cost) and some would stay the same regardless of survey length (“fixed costs” like community announcements, monitoring, and supervision). We thus broke down fixed and variable costs, and we computed the variable costs according to the length of the survey: a shorter survey will likely be less expensive to implement, as enumerators can conduct more surveys each day.

To put these costs into perspective, we compared them with the cost of a full water consumption subsidy for one year. We assumed that the subsidy would consist of providing water for free to identified poor households at a rate of 20 L per person per day, based on the World Health Organization (WHO) recommendation to meet basic needs. We assumed a household size of five (the average in our dataset) and water tariffs of 0.2 GHS (0.03 USD) per 20 L of water based on typical prices at Safe Water Network’s systems.

Finally, to document stakeholder perceptions regarding the different poverty identification methods, we conducted in-depth individual qualitative interviews with six local government officials and 17 community members, including nine individuals designated as poor during community consultation. Interviews focused on perceived fairness and ease of implementation. These qualitative interviews were not meant to give a quantitatively representative picture but rather to capture the full range of opinions. We started these interviews with the goal of capturing a wide-range of individuals (e.g., District Assembly officials occupying different positions, poor and non-poor community members). We stopped interviewing new respondents when we reached saturation, i.e., when themes were recurring and each new interview was bringing little new information (Fusch and Ness, 2015).

2.2.3.3. Models ranking. We compared the five methods with regard to the characteristics of households they identified as poor, scalability, and acceptance among local stakeholders. Then, we conducted a multi-criteria decision analysis (MCDA) to suggest a ranking for different

scenarios that simulated different priorities (Belton and Stewart, 2002; De Luca et al., 2017; Mendoza and Martins, 2006) (Appendix E).

Specifically, we considered the following criteria based on the literature: i) performance: ability to capture households having characteristics associated with greater vulnerability and less wealth, ii) acceptability: perceived fairness and perceived transparency, and iii) scalability: ease of implementation and affordability. For each of the five methods, we assigned a score to these different criteria based on the results from preceding sections, with a score of 1 being the worst and 5 being the best.

2.3. Ethical review

The study protocol was approved by the Council for Scientific and Industrial Research in Ghana (RPN 016/CSIR-IRB/2021).

3. Results

The results section follows the same structure as the previous section. We first assessed the performance of the ML-based PMT at predicting poverty status as defined in the GLSS7 (i.e., relative to the national poverty line) and compared that with the performance of the PPI. Then, we field-tested the five methods to compare them with respect to the characteristics of households they identified as poor, their acceptability by local stakeholders, and their scalability.

3.1. Performance of the ML-based proxy-means test and the PPI at predicting poverty status relative to the national poverty line

The two regional ML-based PMTs that we developed (with 47 and 10 indicators) had an accuracy of 83–87% (when setting the prevalence at 20%), meaning that 83–87% of households were correctly identified as being poor or non-poor (Table 2). The precision of 43–55% meant that about half of households predicted to be poor by our ML-based PMTs were actually poor according to the national poverty line, while the other half was not (inclusion errors). Our PMTs performed better with respect to sensitivity, with 62–73% of truly poor households (i.e., households under the national poverty line) identified as poor (the rest being exclusion errors). Specificity showed the highest performance, with 87–90% of truly non-poor households correctly identified as non-poor (Table 2). The 47-predictor ML-based PMT performed better than the 10-predictor ML-based PMT with respect to all metrics, but particularly precision and sensitivity (Table 2). We therefore focused on the 47-predictor model during subsequent analyses. Both ML-based PMTs performed better than the PPI, which had lower values across all metrics (particularly precision and sensitivity; Table 2). We also computed a random forest algorithm and a multi-linear regression using the 47 indicators as a comparison with the ML-based PMT derived from the metalearning algorithm. The results were very similar (Appendix A, table A.3), but the ML-based PMT performed a bit better, especially for

Table 2

Performance of our two ML-based PMTs and the PPI at predicting poverty status defined according to the national poverty line.

| | ML-based PMT (47 predictors) | ML-based PMT (10 predictors) | PPI (10 predictors) |
|-------------------------------|------------------------------|------------------------------|---------------------|
| Prevalence^a | 20% | 20% | 21% |
| Accuracy | 87% | 83% | 81% |
| Sensitivity | 73% | 62% | 57% |
| Specificity | 90% | 87% | 85% |
| Precision | 55% | 43% | 40% |
| AUC^b | 81% | 74% | 71% |

^a Purposefully set at 20%, a reasonable quota for water subsidy implementation. (The PPI showed a 21% prevalence because a lot of household probabilities were equal to the threshold value).

^b AUC: Area Under the Curve.

the precision metric (55% for the ML-based PMT versus 50% for random forest and 52% for the multi-linear regression). We thus decided to continue focusing on the ML-based PMT for the rest of the analysis.

We also trained a ML-based PMT at national scale (i.e., including the northern regions, but still excluding Greater Accra) to allow for a direct comparison with the PPI, also trained at national scale. We presented the results in Appendix A.4. The national model performed almost as well as the regional one, but included more indicators than the regional model (65 indicators versus 47). While the national dataset is larger (10,072 households in the training set for the national model versus 5925 households for the regional model) and therefore provides better opportunities for machine learning, regional specificities captured in the regional model slightly outweigh this benefit, leading to comparable predictions. Because the regional model had slightly better performance than the national model and included less indications, we decided to continue working with the regional model.

To determine how the households identified by each approach compared with the truly poor as defined by the national poverty line, we summarized economic and water access characteristics for each group (Table 3). We found that households predicted to be poor by the 47-indicator ML-based PMT were more similar to the poor identified by the national poverty line (the reference) than households predicted as poor

Table 3
Characteristics of poor households in test set ($n = 2539$) as determined by i) the national poverty line, ii) our ML-based PMT, and iii) the PPI.

| | Poor households in the test set as defined according to the national poverty line (reference) | Poor households in the test set as determined by our 47-predictor ML-based PMT | Poor households in the test set according to the PPI |
|---|---|--|--|
| Proportion of total population qualifying as “poor” | 15% (380 out of 2539) | 20% (508 out of 2539) | 21% (537 out of 2539) |
| Total annual gross income per adult equivalent (Mean [IQR] ^a) | 10,514 [1270–9806] GHS 1389 [168–1295] USD | 10,640 [1579–11,875] GHS 1406 [209–1569] USD | 17,965 [2441–17,278] GHS 2373 [322–2283] USD |
| Total annual expenditures per adult equivalent (mean [IQR]) | 1201 [952–1502] GHS 159 [126–198] USD | 1785 [1137–2262] GHS 236 [150–299] USD | 2350 [1316–3030] GHS 310 [174–400] USD |
| Percent access to improved water for drinking water | 89% N = 337 | 90% N = 456 | 85% N = 455 |
| Time spent fetching drinking water every day – minutes/day (Mean [IQR]) | 24 [8–30] min/day | 22 [8–30] min/day | 24 [10–30] min/day |
| Daily water consumption (Mean [IQR]) ^a | 17 [6–22] L/capita/day | 17 [6–21] L/capita/day | 18 [5–21] L/capita/day |
| Percent access to piped water | 14% N = 53 | 13% N = 68 | 9% N = 51 |
| Percent access to bottled/sachet water | 23% N = 88 | 24% N = 123 | N = 17% N = 90 |
| Total Water expenditures ^b (Mean [IQR]) | 24 [4–37] GHS/month 3.2 [0.5–4.9] USD/month | 37 [4–45] GHS/month 4.9 [0.5–5.9] USD/month | 30 [5–43] GHS/month 4.0 [0.7–5.7] USD/month |

^a IQR: Interquartile range.

^b Total water expenditures include piped water bills, water purchased from private vendors, water fetched at standpipes, and bottled/sachet water.

by the PPI. Households identified as poor by the PPI had a significantly higher total annual gross income per adult equivalent compared to reference (17,965 GHS vs. 10,514 GHS, $p < 0.05$ Wilcoxon test¹), while the annual income of households identified by the ML-based PMT was not statistically different from reference (10,514 GHS vs 10,640 GHS, $p = 0.05$ Wilcoxon test¹). For both PMTs, total annual expenditures per adult equivalent were significantly higher than reference (1785 GHS vs 1201 GHS for the ML-based PMT; 2350 GHS vs 1201 GHS for the PPI; $p < 0.05$ Wilcoxon test¹ and Z-test). There were no significant differences in water access between the reference and both PMTs, except for households identified as poor by the PPI, who had a lower access to piped water (9% vs. 14%; $p = 0.04$ Z-test), but also a lower access to bottled/sachet water (17% vs. 23%, $p < 0.05$ Z-test) (Table 3).

To understand any incorrect predictions being made by the ML-based PMT, we compared characteristics of households that represented inclusion and exclusion errors with households that were correctly predicted as poor or non-poor (Appendix F – Figure F1. a). With respect to total annual expenditures per adult equivalent, households representing exclusion errors spent slightly more than the correctly predicted poor (1345 GHS vs. 1149 GHS; $p < 0.05$ Wilcoxon test), and households representing inclusion errors (i.e., incorrectly predicted non-poor) tended to spend less than the rest of the non-poor (2559 GHS vs. 5852 GHS; $p < 0.05$ Wilcoxon test). In other words, the errors tended to be close to the threshold between the poor and non-poor. We did the same analysis for the PPI and found similar results (Appendix F.1 – Figure F1. b). We also confirmed that the vulnerability and wealth checks were consistent with true poverty as defined in the GLSS7 (Appendix F.2 – Figure F2).

In summary, the ML-based PMT performed better than the PPI at identifying poor households: households predicted poor by the ML-based PMT tended to be more similar to the poor households identified by the national poverty line in terms of income, expenditures and water characteristics than the households predicted poor by the PPI.

3.2. Field comparison

3.2.1. Characteristics of surveyed households

Overall, we conducted a total of 818 surveys; this sample size allowed estimating the proportion of poor households with a margin of error of 3% and 90% confidence. We excluded 63 additional households in target neighborhoods from the survey: 50 because they were absent over the course of three visits, one because the respondent declined to participate, and 12 because we had concerns about data quality. Our 818 surveys thus covered 93% of all (881) households in the selected neighborhoods.

Heads of household made up 68% of respondents, with 59% of all household heads being males. Socio-economic and demographic characteristics of the surveyed households are displayed in Table 4, and we compared them with the households included in the regional GLSS7 dataset. Approximately half of household heads (54%) worked in agriculture, and 55% completed high school. In 53% of households, the head was either a single woman, an elderly person, or a person with a severe physical or mental disability or chronic illness, and 22% of the households reported not being able to feed themselves throughout the entire year without help from relatives. Only a small proportion of the households used unimproved or surface water (1%), and the vast majority (85%) used either piped water (through communal standpipes, 52%) or other improved sources such as boreholes (33%). The rest (15%) used sachet or bottled water. Notably, <1% household had a private connection to the piped network, as water suppliers were not yet offering this modality. These socio-economic and demographic characteristics of the surveyed households were almost all within a few percentage points of the regional GLSS7 dataset, with a few exceptions: surveyed households usually owned farming land or livestock more

¹ Wilcoxon rank sum test for independent samples.

Table 4

Demographic and socio-economic characteristics of surveyed households (N = 818) and southern Ghana households included in the GLSS7 dataset (N = 8464).

| Characteristics of surveyed households within the three targeted villages (N = 818) | | Characteristics of households included in the Southern Ghana GLSS7 dataset (N = 8464) | p-value |
|---|-------------|---|-----------------------|
| Mean number of household members (IQR) ^a | 5 (3–6) | 3.7 (2–5) | p < 0.05 ^f |
| Mean age of household head (IQR) ^a | 48 (35–58) | 46 (33–57) | p < 0.05 ^f |
| Gender of household head | | | |
| Female | 41% N = 337 | 36% N = 3008 | p < 0.05 ^g |
| Male | 59% N = 481 | 64% N = 5456 | p < 0.05 ^g |
| Education level of household head | | | |
| Never attended school | 25% N = 206 | 17% N = 1462 | p < 0.05 ^g |
| No primary education | 7% N = 58 | <1% N = 8 | p < 0.05 ^g |
| Completed only primary school | 13% N = 107 | 16% N = 1346 | p = 0.03 ^g |
| Completed high school or higher ^b | 55% N = 447 | 66% N = 5606 | p < 0.05 ^g |
| Marital status of household head | | | |
| Married or in a union | 59% N = 479 | 60% N = 5081 | p = 0.43 ^g |
| Single or separated/divorced | 21% N = 178 | 27% N = 2325 | p < 0.05 ^g |
| Widowed | 19% N = 159 | 12% N = 1058 | p < 0.05 ^g |
| Primary occupation of household head | | | |
| No occupation | 16% N = 129 | 12% N = 1041 | p < 0.05 ^g |
| Agriculture | 54% N = 445 | 38% N = 3222 | p < 0.05 ^g |
| Other occupation | 30% N = 244 | 50% N = 4201 | p < 0.05 ^g |
| Primary religion of household head | | | |
| Muslim | 17% N = 136 | 10% N = 864 | p < 0.05 ^g |
| Christian | 82% N = 668 | 12% N = 993 (Catholic) | p < 0.05 ^g |
| Head of household is vulnerable | | | |
| Single woman/widow/separated | 32% N = 264 | 20% N = 1706 | p < 0.05 ^g |
| Elderly (65 years or older) | 16% N = 132 | | |
| Severe physical/mental challenge ^c | 3% N = 24 | 14% N = 1196 | p = 0.13 ^g |
| Severe chronic illness ^c | 2% N = 15 | 3% N = 230 (any disability) | p < 0.05 ^g |
| Household is a beneficiary of the LEAP program | 4% N = 31 | <1% N = 68 | p < 0.05 ^g |
| Household owns farming land | 31% N = 251 | 15% N = 1297 | p < 0.05 ^g |
| Household owns livestock | 44% N = 360 | 9% N = 770 | p < 0.05 ^g |
| Households that can't feed themselves | 22% N = 178 | 58% N = 4924 | p < 0.05 ^g |
| Households with a non-durable dwelling ^d | 15% N = 121 | 46% N = 3907 | p < 0.05 ^g |
| Gross annual income ^e | | | |
| <2400 GHS | 54% N = 443 | 25% N = 2117 | p < 0.05 ^g |
| 2400–6000 GHS | 29% N = 241 | 19% N = 1616 | p < 0.05 ^g |
| >6000 GHS | 17% N = 134 | 56% N = 4731 | p < 0.05 ^g |
| Time to fetch drinking water, min (IQR) | 9 (3–10) | 22 (7–30) | p < 0.05 ^f |
| Primary source of drinking water | | | |
| Bottled or sachet water | 15% N = 123 | 24% N = 2065 | p < 0.05 ^g |

Table 4 (continued)

| Characteristics of surveyed households within the three targeted villages (N = 818) | | Characteristics of households included in the Southern Ghana GLSS7 dataset (N = 8464) | p-value |
|---|---------------|---|-----------------------|
| Piped water (communal standpipes) | 52% N = 420 | 11% N = 963 | p < 0.05 ^g |
| Piped water (private connections) | <1% N = 1 | 14% N = 1191 | p < 0.05 ^g |
| Improved, non-piped (boreholes, protected spring and well) | 33% N = 271 | 39% N = 3291 | p < 0.05 ^g |
| Unimproved | <1% N = 1 | | |
| Surface water | <1% N = 2 | 11% N = 953 (unimproved and surface water) | p < 0.05 ^g |
| Mean weekly water expenditures GHS (IQR) ^a | 4.5 [1.6–7.0] | 8.8 (2.1–9) | p < 0.05 ^f |

^a IQR = inter-quartile range.

^b Preventing head of household from working.

^c High school included junior and senior high school, A level, and teacher-nursing-technical training.

^d Non-durable is defined as a dwelling with natural or rudimentary roof and walls (Ghana Statistical Service (GSS) et al., 2018).

^e Five households replied “Don’t know” to this question.

^f Wilcoxon test.

^g Z-test.

frequently than the households in the GLSS7 dataset (31% and 44% versus 15% and 9%); a lesser percentage reported not being able to feed themselves all year round compared to the GLSS7 dataset (22% versus 58%); access to piped water was higher among the surveyed households (52% versus 25%), while the percentage of households using unimproved or surface water was higher in the GLSS7 dataset (<1% versus 11%). The three towns are within the service areas of SWN, which might have increased access to piped water services in comparison to other parts of Southern Ghana. Overall, except for the characteristics mentioned above, the surveyed towns were a good representation of Southern Ghana households.

3.2.2. Characteristics of households identified as poor by the five targeting methods

The three PMTs were set to identify 20% of households as poor: 164/818 for ML-based PMT and DHS and 169/818 for PPI, corresponding to households within the first quintile of poverty probability or wealth index. The other two methods were more restrictive: only 4% of households (31/818) reported they were enrolled in the LEAP program, and CBT identified 12% (98/818) of households as poor. There was limited overlap in the households identified as poor across the five methods. Only two households (<1%) were identified as poor by all five methods, and 10 households (1%) were identified as poor by four methods (excluding LEAP, the most restrictive method). In contrast, there was more overlap among the three PMT methods: 65/818 households (8% of the total number of households) were identified simultaneously as poor by all three PMT methods. These 65 households represented 40% of all those identified as poor by any one of the three PMTs. These results reflect that each method used different criteria to define poverty, but the PMT methods seemed to be more consistent with each other.

All methods successfully identified poor households that were more vulnerable than the rest, according to our checks: households that were selected as poor met the different vulnerability checks more often than other households not selected (Fig. 3a). For example, 30%–67% of households identified as poor met at least two vulnerability checks versus 12%–21% for households identified as non-poor. With the exception of LEAP, all the methods identified non-poor households that

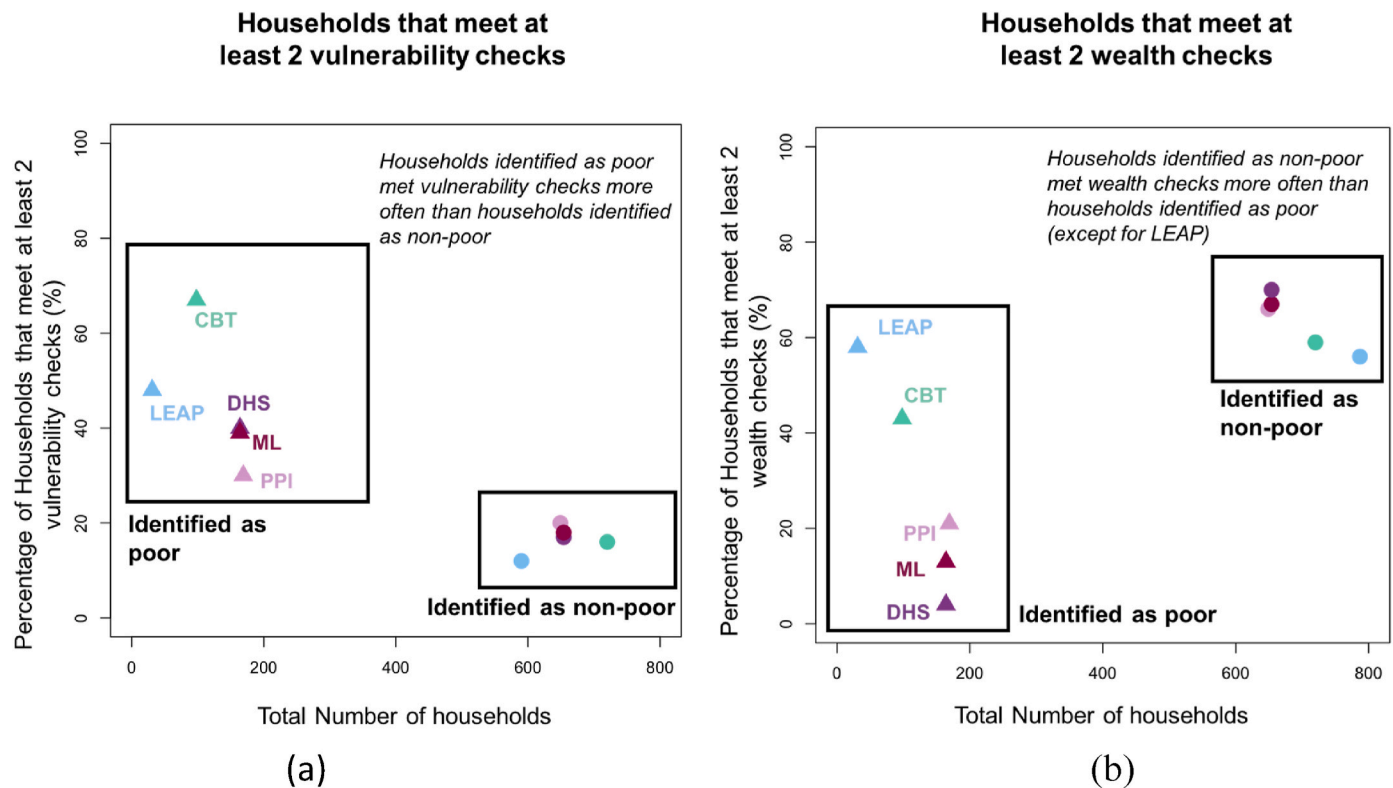


Fig. 3. Percentage of households identified as poor and non-poor by the methods and meeting at least two vulnerability checks (a) or two wealth checks (b) versus the total number of households identified as poor or non-poor.

met wealth checks more often than households identified as poor: 58%–70% of non-poor households met at least two wealth criteria versus 4%–43% of households identified as poor. In contrast, LEAP identified the same proportion of poor and non-poor households meeting the wealth checks (58% of identified poor households versus 57% of non-poor households - Fig. 3b). Numerical results are presented in Appendix G - Table G.1.

According to our vulnerability checks, CBT had the highest ability to identify households more vulnerable than the rest: it provided the highest vulnerability gap between households identified as poor and non-poor (51% versus 10–23% for other methods, Fig. 3a and Table G1). This is not surprising, as the guidance facilitators provided during CBT was similar to our vulnerability checks (Appendix B). With respect to wealth checks, the DHS PMT provided the highest gap between households identified as poor and non-poor (82%) while the other two PMTs yielded smaller gaps (45–54%) (Fig. 3b and Table G1): these results are consistent with the fact that the DHS wealth index was based on asset ownership, while the ML-based PMT and the PPI were based on consumption data. LEAP and CBT led to negligible wealth gaps (–1% and 15%), reflecting that these methods rely on community perceptions of poverty that do not seem to take asset ownership into account.

We also looked at the characteristics of households left behind by one method but identified as poor by another one. The main characteristic of these households were: a female household head and inability to feed all year round (Appendix I).

3.2.3. Water access conditions

As the goal of this targeting exercise was to identify poor households for water subsidies, we explored water needs among households identified as poor by the five methods and also between poor and non-poor (Figs. 4 and 5). We wanted to determine the extent to which households identified as poor by the methods were also the most in need of safe water.

Households identified as poor by any method had lower use of

packaged (sachet or bottled) water than the non-poor ($p < 0.05$ Z-test, except for LEAP where $p = 0.27$ - Fig. 4a) mirrored by a higher reliance on improved non-piped water sources ($p < 0.05$ Z-test, except for CBT and LEAP: $p = 0.36$ and $p = 0.14$). The three PMTs yielded similar proportions of standpipe water users among the poor and non-poor, while CBT and LEAP yielded higher proportion of standpipe water users among the poor than among the non-poor (Fig. 4a). Note that as mentioned above, no household in our study communities had private piped water connections.

Across all methods except LEAP, households identified as poor reported financial constraints to access water more often (22–33%) than the non-poor (5–10%) ($p < 0.05$ Z-test - Fig. 4b). 13% of households identified as poor by CBT reported a physical inability to fetch water, compared to <4% among other categories. This is consistent with the CBT using physical impairment (due to disability or old age) as one of the guiding criteria.

Regarding water consumption per person per day (Fig. 5a), households identified to be poor by the three PMTs had lower levels of consumption than the non-poor ($p < 0.05$ Wilcoxon test), but there was no statistical difference between households identified as poor or non-poor by the CBT and LEAP ($p = 0.27$ and $p = 0.48$ Wilcoxon test). Among poor households, the PPI and the ML-based PMT more effectively identified poor households with lower water consumption ($p < 0.05$ Wilcoxon test with DHS and CBT), and there was no statistical difference between the PPI and the ML-based PMT ($p = 0.25$ Wilcoxon test).

We found no differences in the time to fetch water between households identified as poor and non-poor by all the methods ($p > 0.05$ Wilcoxon test, except for CBT where $p = 0.02$ but numerical differences were negligible) (Fig. 5b).

Among the five methods, the PMTs were better able to identify households with greater water needs (i.e., no use of packaged water, more frequent reporting of financial burden to get water, and lower water consumption). We note that 3–7% of households identified as non-poor by the methods still reported financial constraints to get water

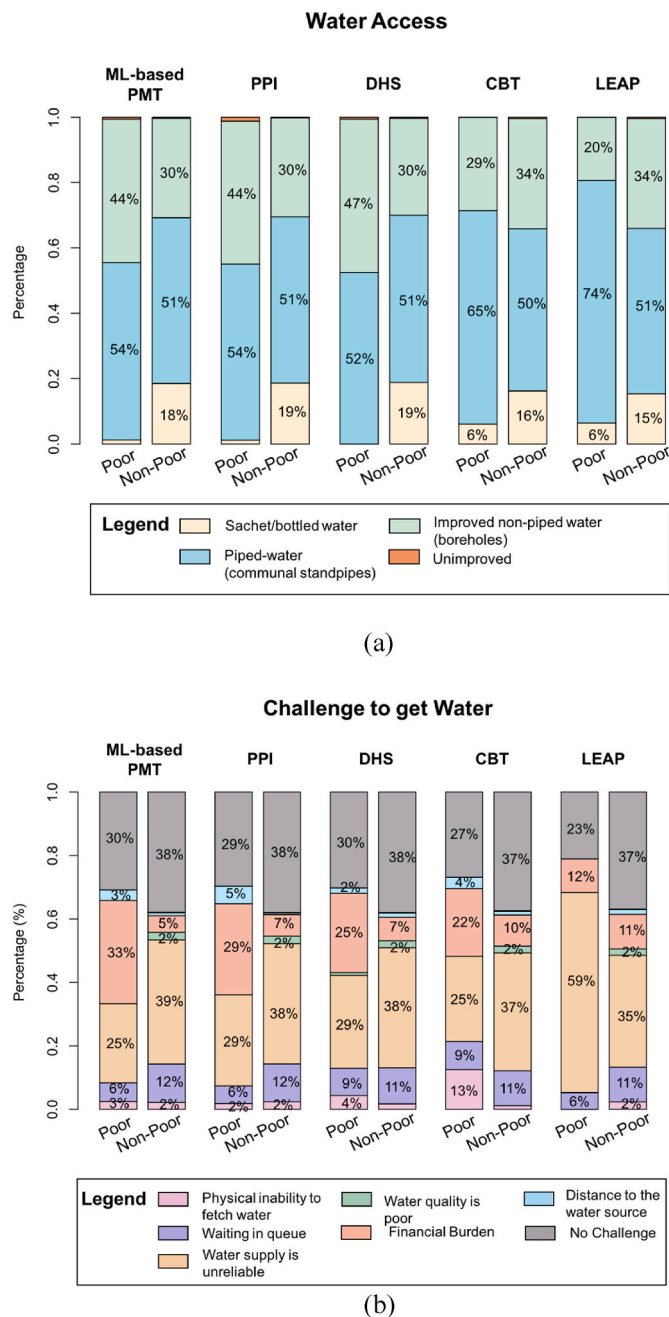


Fig. 4. Primary source of drinking water among poor and non-poor households as determined by each of the five identification methods (a) and challenges to get water reported by households identified as poor and non-poor by the five different methods (b).

(Appendix H – Figure H.1). The principal characteristics of these households were that the household had a female head (62–78% of households left behind) or that the household cannot feed without help from relatives all year round (29–34%) (Appendix H – Figure H.2).

3.2.4. Stakeholder and community member perceptions

During qualitative stakeholder interviews, we referred to the three PMT methods together as “survey-based approaches” for simplicity. Community members were overwhelmingly in support of survey-based approaches to target water subsidies. They felt that people were more likely to be honest and forthcoming in a household survey, whereas they may be afraid to share their true poverty status in community meetings. Interviewees also felt that CBT might leave vulnerable members off the

list because people may not be aware of others’ true poverty status and vulnerable households may not be able to attend the meeting. “One might be poor and your neighbors don’t know you are poor until you tell them,” stated a respondent. “If, unfortunately, the poor and vulnerable person was not remembered by any of the participants at the meeting, then he or she was left out ... One of my relatives was left out of the list. In fact, I forgot of her, it was after the meeting I remembered of her,” stated one of the meeting participants.

In contrast, District Assembly officers had mixed perspectives about community consultation and survey-based methods: some preferred CBT, while others preferred survey-based methods. Most of them feared that people might lie during the household survey. “If I’m alone, I feel comfortable lying. But if there are others around, sometimes you have a second thought of lying,” said one District Assembly officer. Some District Assembly officers felt that people were less likely to lie in the community consultation meeting because of the process of vetting poor households collectively, though others feared that meeting participants might lie anyway. Finally, some District Assembly officers felt that CBT would reduce accusations of political bias against the District Assembly because it is the most transparent targeting method. “Using that [community consultation] approach, it’s more or less like a fair interaction between the District Assembly and the community. There will be no perception that the Assembly was biased or selective in choosing people to give those subsidies to,” said one District Assembly officer.

Community members overwhelmingly felt that LEAP would not be an appropriate method to target water subsidies, because they felt that the beneficiary selection was influenced by politics and that the risk of leaving vulnerable community members off the list was high (they felt that LEAP would discriminate against people and would fail to target all the poor). “If they were to use the LEAP list, some of us will be left out” stated one community member identified as poor in the community consultation. Similarly, no District Assembly officers chose LEAP as their preferred targeting method. Most District Assembly officers thought that LEAP would lead to accusations of political bias. “The way the previous [LEAP] registration was done, it didn’t really capture the vulnerable and the poor in the communities” stated one District Assembly officer.

3.2.5. Costs

Based on our cost tracking, implementing PMT surveys across our six study neighborhoods cost 4558 USD, or an average of 5.6 USD per household in the community (N = 818 households) which was nearly twice as much as CBT (2474 USD or 3.0 USD per household in the community; Appendix J). Survey-based methods were costlier because they required substantially more staff time: 21 person-days per neighborhood on average to survey all households, compared with 9 person-days for CBT. The survey methods included fixed costs (26% of the total) and variable costs (74% – see Appendix J – Figure J.1). Fixed costs will remain the same regardless of the length of the survey; they will change only if additional neighborhoods or towns are to be surveyed. Fixed costs included enumerators training, transport to communities for community entry and community announcements, facilitation allowances to chiefs and traditional leaders, chair rental, phones and communication allowances, and COVID-19 personal protective equipment. Variable costs will change depending on the length of the survey: a shorter survey will reduce staff time (and thus staff payments), some of the transportation costs for administering the survey, and the communication allowance. An estimation of the variable costs based on the time needed for the survey to be completed is provided in Appendix J.

Finally, we estimated that determining subsidy-eligible households using the existing LEAP lists would only cost 533 USD, or an average of 0.7 USD per household in the community as the staff time requirements would be minimal (3 person-days per neighborhood).

Our cost estimates for the water subsidy itself tended to be substantially higher than those for targeting (and this was a conservative comparison, as these subsidy costs only included the cost of the water without any program implementation or administrative costs). One year

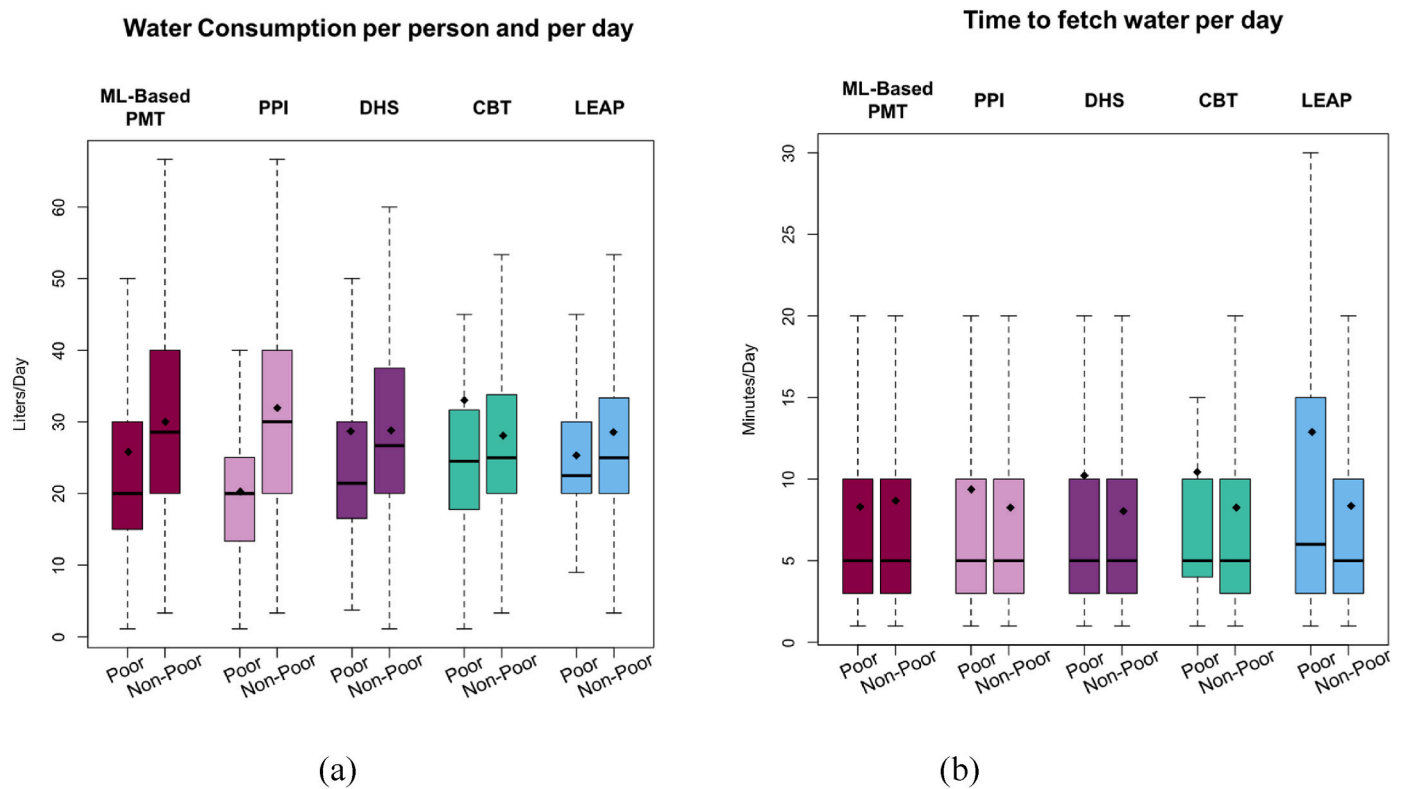


Fig. 5. Water consumption per person per day among households identified as poor and non-poor by the five methods (a) and time to fetch water among households identified as poor and non-poor by the five different methods (b). Whisker plot: the box represents the first and the third interquartile, the black horizontal bar is the median and the black dot represents the mean.

of water subsidies was most expensive for the PMT models, as they were the most inclusive methods, while water subsidies for LEAP were the cheapest, as it identified only 4% of households as poor (Appendix J – Figure J.1). One year of water subsidies cost between 2.3 times (PMTs) and 3.9 times (LEAP) the cost of targeting, depending on the method and the number of households it selected (Appendix K).

3.2.6. Methods ranking

For the methods ranking, we converted our field results regarding household characteristics, acceptance, and scalability into scores for each targeting method (Table 5). With respect to household characteristics, we ranked the methods according to their ability to identify households that were more vulnerable than the rest (according to vulnerability and wealth checks). For the ML-based PMT and the PPI,

Table 5
Scores assigned scores to five poverty targeting methods according to six comparison criteria.

| Score (Best = 5, worst = 1) | ML-based PMT | PPI | DHS | CBT | LEAP |
|--|--------------|-----|-----|-----|------|
| Characteristics of households identified | | | | | |
| Few exclusion and inclusion errors | 4 | 3 | NA | NA | NA |
| Identified households are more vulnerable than the rest ^a | 4 | 2 | 5 | 3 | 1 |
| Acceptability | | | | | |
| Perceived fairness | 5 | 5 | 5 | 3 | 1 |
| Perceived transparency | 4 | 4 | 4 | 5 | 1 |
| Scalability | | | | | |
| Ease of implementation | 1 | 3 | 2 | 4 | 5 |
| Targeting affordability (=Low cost) | 2 | 2 | 2 | 4 | 5 |

^a To assign scores, we summed the vulnerability gap and the wealth gap (i.e., the % difference between poor and non-poor in Table G1) for each method. The method with the highest sum gap (DHS) ranked the highest.

which we could directly evaluate in the GLSS7 dataset against the reference poverty definition, we also considered inclusion and exclusion errors. For acceptance, we referred to the stakeholder and community member interviews to rank the methods. Finally, for scalability, we assigned scores based on our cost computations and our understanding of what each method would require for implementation at a large scale.

With respect to inclusion/exclusion errors (i.e., performance), we attributed a higher score to the ML-based PMT than the PPI based on the results reported in Table 2. The DHS had the highest ability to identify households that were more vulnerable and less wealthy than the rest (according to our vulnerability and wealth checks), followed by the ML-based PMT, CBT, the PPI, and LEAP. We determined these rankings based on vulnerability/wealth gaps between households identified as poor and non-poor (Fig. 3 and Table G1).

We gave the PMTs the highest “fairness” score, as the majority of community members interviewed as well as some District Assembly officers perceived survey approaches to be fairer than the CBT. We then attributed the highest “transparency” scores to CBT followed by the PMTs because of their “black-box” approach. We ranked LEAP the last because no interviewees perceived it to be fair or transparent.

We gave LEAP the best implementation score because LEAP beneficiaries had already been designated by the government, and the lists were available from government offices. CBT is in the second position as it does not require interviewing all the households in a community. The PMT models came in the last position: they would require more logistical considerations, enumerators, and training. Among the PMT methods, the PPI is the easiest to implement: it contains only 10 questions, and the results of those questions can easily be turned into a poverty score using a scorecard and lookup table (PPI, 2019), so it does not require intensive computing. The DHS wealth index involves more questions but, like the PPI, does not require complex computing as the equation to compute a households’ index value is publicly available (Rutstein, 2008; Rutstein and Johnson, 2004). Finally, the ML-based PMT model has the lowest

score as it requires someone knowledgeable with computational techniques to apply the ML model to survey responses.

4. Discussion – conclusion

To improve safe water access, we compared five different methods to target poor households eligible for water subsidies in southwestern Ghana. We first designed a regional ML-based PMT by applying machine learning techniques to the GLSS7 dataset, which performed better than an existing PMT (the PPI) at predicting poverty status defined according to the national poverty line (the reference). We also compared the results of the metalearning algorithm used to develop the ML-based model with two other algorithms (random forest and multi-linear regression). The results of the multi-linear regression were very similar to those derived from the metalearning algorithm, which is consistent with previous work (Christodoulou et al., 2019). For future work, we would recommend considering regional ML-based PMTs rather than national ones if the size of the regional dataset is substantially higher than the national dataset (in the case of local intense survey initiatives for example). If the national and regional dataset have the same size (or if the regional dataset is smaller than the national one), we would recommend to test both models (one at regional scale and one at national scale), as the benefits of regional specificity can outweigh the benefits of training the ML model on a larger dataset.

We then compared the five methods for identifying the poor in the field. The methods resulted in vastly different outcomes, both in terms of the households identified and the overall proportion. The very small overlap between households identified as poor by CBT and by the PMT methods revealed either a different poverty concept between communities and the statisticians developing PMTs, or the possibility that communities do not always understand each other's condition. Previous studies have shown that communities used multidimensional concepts of poverty that include health, education, and household demographic characteristics (e.g., elderly, widow) (Van Campenhout, 2007; Alatas et al., 2012; Hillebrecht et al., 2020), and CBT has been found to be more effective than PMT methods at identifying chronic poverty (Alderman, 2002), defined as significant deprivations experienced by a household for a period of five years or more (Hulme and Shepherd, 2003).

According to the ranking analysis, the best model was the ML-based PMT, which was particularly strong in terms of performance and acceptance, followed by CBT, which represented a good tradeoff between performance, acceptability, and scalability. The simplicity of eligibility criteria in CBT (and the fact that communities were directly involved in household selection and fully understood the process) is also an advantage over survey-based approaches, which rely on “black-box” models to derive a poverty probability score. However, consultative meetings could be subject to ethnic favoritism (Hillebrecht et al., 2020) or to political divides along party lines (Panda, 2015), even though the latter does not seem to impact CBT results substantially (Alatas et al., 2019) and none of these biases were mentioned in our stakeholder interviews. The success of both consultative meetings and household surveys also depends on skilled facilitators, which means there is a risk of poor-quality data or poor implementation affecting identification. Consultative meetings are risky: for example, if the meeting is dominated by a small group of individuals or subject to low attendance due to poor community entry, then there is no clear recourse to correct the list of identified households. In terms of cost, we found that the targeting cost for CBT was lower than for the PMTs, even though local stakeholders did not report cost as a primary factor for determining the most appropriate method.

Our results confirmed the conclusions of the few previous efforts made to compare PMTs with CBT: PMT typically shows better performance than CBT (Alatas et al., 2012; Hillebrecht et al., 2020), while CBT is usually more cost-effective (except for large programs, where the accuracy of PMT makes this method more cost-effective due to reductions in inclusion errors) (Hillebrecht et al., 2020). CBT also typically

leads to higher satisfaction among communities (Alatas et al., 2012).

In scenarios where decision-makers highly value accuracy, a PMT (especially the ML-based PMT) would be the best fit among the options we considered here. For example, in Ghana, the Community Water and Sanitation Agency (CWSA, responsible for delivering safe water and sanitation services to rural areas) indicated that accuracy (i.e., low risk of inclusion and exclusion errors) was their highest priority.

To reconcile performance (ML-based PMT) and transparency (CBT), a hybrid approach could be envisaged. Previous research and programs (such as LEAP) used such a hybrid approach to identify poor households: typically consisting of first asking a community who is poor (CBT), and then surveying the designated households with a PMT (Alatas et al., 2012; Stoeffler et al., 2016). This type of hybrid method is cost-effective (Stoeffler et al., 2016), but has lower performance than a PMT alone, due to a higher risk of exclusion errors (because not all households are surveyed) (Alatas et al., 2012; Stoeffler et al., 2016). To increase transparency without decreasing the performance, an alternative might involve conducting both methods in parallel (PMT and CBT), and then adding the poor households identified by CBT to the list of poor households identified by the PMT (in other words, selecting households identified by either method – see Appendix L), or conducting the PMT first and then having the results discussed and confirmed by the community. This hybrid approach would increase the cost of targeting but may provide a good balance between performance and transparency. We recommend that future research evaluates the hybrid approach of conducting PMT and CBT in parallel, and selecting households identified by either method.

The relatively good performance of the ML-based PMT was balanced by a few limitations. This PMT was developed for Ghana only (one model was developed for Southern Ghana only and another for the entire country) and cannot be directly applied to other countries. However, the methodology (feature reduction and predictive modelling) can be replicated to other countries if large-scale household surveys (covering topics including household characteristics, consumption, and assets) are available. Such replication would be similar to the way in which the PPI and DHS methods have been applied across multiple countries. In our case, however, comparing the ML-based PMT and DHS using the GLSS7 dataset was not straightforward, as we could not directly compare the DHS with the reference (some indicators included in the DHS survey were absent from the GLSS7 dataset). In addition, because we did not collect the large amount of economic data needed to calculate consumption and determine true poverty in the field, we were unable to perform a true out-of-sample evaluation. Instead, we used the vulnerability and wealth checks to determine whether local stakeholders would likely perceive that households identified as poor were more vulnerable than the rest. This analysis provided an additional window into the acceptability of the methods, as a targeting method would have difficulty being accepted if its results were wholly different from local perceptions. Finally, the “black box” approach of the ML-based PMT makes this method difficult for non-data scientists to replicate, but designing a simple web interface that computes poverty predictions based on answers to the 47 survey questions can be envisaged to facilitate the deployment of this method in Ghana. A similar approach could be used in other places, after the structure of the model has been determined through machine learning.

Our data allowed us to examine the implications of targeting based on poverty status rather than water access (as is currently suggested in the Government of Ghana guidelines). While all targeting methods (particularly the three PMTs) identified households with greater water needs than the rest of the population, there was only partial overlap between those identified as poor and those not currently accessing piped water. In practice, water suppliers may find it difficult to allocate subsidies to households who already have access and/or to leave out households without access. If that is the case, targeting methods that consider both poverty status and water access could be conceived. These decisions may also vary based on the type of water access that the

supplier is hoping to promote. In our three study communities, 52% of households already accessed piped water through communal standpipes, but none had private piped connections. If the water supplier's goal is to promote private connections, it may use poverty status alone to determine who is eligible for subsidies. In contrast, if the water supplier is not planning to introduce private connections and instead wants to further promote standpipe use, it may decide to apply poverty screening to non-customers only, as opposed to the general population.

By highlighting the strengths and the weaknesses of five different targeting methods (new ML-based PMT, two existing PMTs, LEAP, and CBT), this study can provide guidance to practitioners in choosing the most appropriate method to target poor households eligible for water subsidies, considering performance, cost, and acceptability by local stakeholders. Appropriate targeting of water subsidies can help make access to safe water more equitable and thereby improve public health. Beyond targeting, further research should examine approaches for administering and financing water subsidies without threatening the financial stability of service providers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was funded through a grant from the Conrad N. Hilton Foundation to the Aquaya Institute [Grant number 17284]. We thank Lauren Cuscuna, Joseph Ampadu-Boakye, Francis Tetteh-Zomayi and Emmanuel Anning from Safe Water Network for their help in liaising with District Assemblies and communities for the field evaluation, as well as IRC Ghana for facilitating consultations with District Assemblies. We also want to express our gratitude to the District Assemblies of Asutifi North, Tano South, and Ahafo Ano North for their active participation in our field evaluation. Additionally, we thank the Asutifi North District Assembly and the Community Water and Sanitation Agency (CWSA) for providing inputs on adequate targeted subsidy designs. Finally, this work was only possible thanks to our enumerators Daniel Nfoke, Moses Kumba, and Robert Kampi Laari, to the cooperation of the community leaders and to the hundreds of community members who generously donated their time for our survey and community meetings.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.deveng.2022.100098>.

References

- Abramovsky, L., Andres, L., Joseph, G., Rud, J.P., Sember, G., Thibert, M., 2020. Study of the Distributional Performance of Piped Water Consumption Subsidies in 10 Developing Countries. World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-9245>.
- Ahmad, M.W., Reynolds, J., Rezgui, Y., 2018. Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* 203, 810–821. <https://doi.org/10.1016/j.jclepro.2018.08.207>.
- Aiken, E., Bellue, S., Blumenstock, J., Karlan, D., Udry, C., 2021a. Using mobile phone and satellite data to target emergency cash transfers in Togo. URL: <https://www.poverty-action.org/study/using-mobile-phone-and-satellite-data-target-emergency-cash-transfers-togo>, 9.6.21.
- Aiken, E., Bellue, S., Karlan, D., Udry, C.R., Blumenstock, J., 2021b. Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance (Working Paper No. 29070), Working Paper Series. National Bureau of Economic Research. <https://doi.org/10.3386/w29070>.
- Aiken, E.L., Bedoya, G., Coville, A., Blumenstock, J.E., 2020. Targeting development aid with machine learning and mobile phone data: evidence from an anti-poverty intervention in Afghanistan. In: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies. COMPASS '20. Association for Computing Machinery, New York, NY, USA, pp. 310–311. <https://doi.org/10.1145/3378393.3402274>.
- Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Purnamasari, R., Wai-Poi, M., 2019. Does elite capture matter? Local elites and targeted welfare programs in Indonesia. *AEA Papers Proc.* 109, 334–339. <https://doi.org/10.1257/pandp.20191047>.
- Akaligaung, Akalpa J., Barrington, Clare, Sudhanshu, Handa, Oosei-Akotoo, Isaac, 2016. Ghana LEAP 1000 programme: baseline evaluation report. Republic of Ghana - Ministry of Gender 1, 1–88. *Children and Social Protection*.
- Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Purnamasari, R., Wai-Poi, M., 2016. Self-targeting: evidence from a field experiment in Indonesia. *J. Polit. Econ.* 124, 371–427. <https://doi.org/10.1086/685299>.
- Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Tobias, J., 2012. Targeting the poor: evidence from a field experiment in Indonesia. *Am. Econ. Rev.* 102, 1206–1240. <https://doi.org/10.1257/aer.102.4.1206>.
- Alderman, H., 2002. Do local officials know something we don't? Decentralization of targeted transfers in Albania. *J. Publ. Econ.* 83, 375–404. [https://doi.org/10.1016/S0047-2727\(00\)00145-6](https://doi.org/10.1016/S0047-2727(00)00145-6).
- Andres, L.A., Thibert, M., Lombana Cordoba, C., Danilenko, A.V., Joseph, G., Borja-Vega, C., 2019. Doing More with Less: Smarter Subsidies for Water Supply and Sanitation, Water Papers. World Bank. <https://doi.org/10.1596/32277>.
- Aryeetey, G.C., Jehu-Appiah, C., Spaan, E., D'Exelle, B., Agyepong, I., Baltussen, R., 2010. Identification of poor households for premium exemptions in Ghana's National Health Insurance Scheme: empirical analysis of three strategies. *Trop. Med. Int. Health* 15, 1544–1552. <https://doi.org/10.1111/j.1365-3156.2010.02663.x>.
- Ashitey, J., Yeboah, S.A., 2012. Livelihood Empowerment against Poverty Program (LEAP) - Operations Manual.
- Belton, V., Stewart, T., 2002. Multiple Criteria Decision Analysis: an Integrated Approach. Springer Science & Business Media.
- Brown, C., Ravallion, M., van de Walle, D., 2018. A poor means test? Econometric targeting in Africa. *J. Dev. Econ.* 134, 109–124. <https://doi.org/10.1016/j.jdevco.2018.05.004>.
- Chatterjee, S., Simonoff, J.S., 2013. Handbook of Regression Analysis. Wiley Online Library.
- Cheng, L.-H., Hsu, T.-C., Lin, C., 2021. Integrating ensemble systems biology feature selection and bimodal deep neural network for breast cancer prognosis prediction. *Sci. Rep.* 11, 14914 <https://doi.org/10.1038/s41598-021-92864-y>.
- Chi, G., Fang, H., Chatterjee, S., Blumenstock, J.E., 2021. Micro-Estimates of Wealth for All Low-And Middle-Income Countries arXiv preprint arXiv:2104.07761.
- Christodoulou, E., Ma, J., Collins, G.S., Steyerberg, E.W., Verbakel, J.Y., Van Calster, B., 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* 110, 12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>.
- De Luca, A.L., Iofrida, N., Leskinen, P., Stillitano, T., Falcone, G., Strano, A., Gulisano, G., 2017. Life cycle tools combined with multi-criteria and participatory methods for agricultural sustainability: insights from a systematic and critical review. *Sci. Total Environ.* 595, 352–370. <https://doi.org/10.1016/j.scitotenv.2017.03.284>.
- Dershem, L., Saidulloev, F., Nadareishvili, M., Arnold, C., Rittmann, J., 2013. Using a Proxy Means Test for Targeting in a Conditional Cash Transfer Program, BOTA Foundation, IREX and Save the Children.
- Foster, V., Gómez-Lobo, A., Halpern, J., 2000. Designing direct subsidies for the poor - a water and sanitation. *Case Study* 211, 4.
- Fosu, K., 2020. Ghana to Provide 550,000 People with Improved Water and Sanitation Services. World Bank.
- Fusch, P.I., Ness, L.R., 2015. Are we there yet? Data Saturat. *Qualitative Res.* 9.
- Ghana, 2018. Statistical Service (GSS), Ghana Health Service (GHS), ICF. Ghana Maternal Health Survey 2017. Accra, Ghana: GSS, GHS and ICF.
- GLSS7, 2018. Poverty Trends in Ghana 2005-2017. Ghana Living Standard Survey.
- Handa, S., Park, M., 2012. Livelihood Empowerment against Poverty Program Ghana Baseline Report, vol. 37. The university of North Carolina at Chapel Hill.
- Hargreaves, J.R., Morison, L.A., Gear, J.S., Kim, J.C., Makhubele, M.B., Porter, J.D., Watts, C., Pronyk, P.M., 2007. Assessing household wealth in health studies in developing countries: a comparison of participatory wealth ranking and survey techniques from rural South Africa. *Emerg. Themes Epidemiol.* 4, 4. <https://doi.org/10.1186/1742-7622-4-4>.
- Hillebrecht, M., Klöner, S., A Pacere, N., Souares, A., 2020. Community-based versus statistical targeting of anti-poverty programs: evidence from Burkina Faso. *J. Afr. Econ.* 29, 271–305. <https://doi.org/10.1093/jafeco/ejz026>.
- Hoo, Y.R., Joseph, G., Rivera, R., Smets, S., Nguyen, H., Ljung, P., Um, S., Davis, G., Albert, J., 2022. Strategic complements: poverty-targeted subsidy programs show additive benefits on household toilet purchases in rural Cambodia when coupled with sanitation marketing. *PLoS One* 17, e0269980. <https://doi.org/10.1371/journal.pone.0269980>.
- Hulme, D., Shepherd, A., 2003. Conceptualizing chronic poverty. *World Dev.* 31, 403–423. [https://doi.org/10.1016/S0305-750X\(02\)00222-X](https://doi.org/10.1016/S0305-750X(02)00222-X).
- Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 790–794. <https://doi.org/10.1126/science.aaf7894>.
- Jmp, 2019. Progress on Household Drinking Water, Sanitation and Hygiene 2000-2017. Special Focus on Inequalities. United Nations Children's Fund (UNICEF) and World Health Organization (WHO), New York.

- Karasu, S., Altan, A., Bekiros, S., Ahmad, W., 2020. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy* 212, 118750. <https://doi.org/10.1016/j.energy.2020.118750>.
- Kshirsagar, V., Wiecezorek, J., Ramanathan, S., Wells, R., 2017. Household Poverty Classification in Data-Scarce Environments: a Machine Learning Approach arXiv: 1711.06813 [stat].
- Laan, M.J. van der, Polley, E.C., Hubbard, A.E., 2007. Super Learner. *Statistical Appl. Genet. Mol. Biol.* 6 <https://doi.org/10.2202/1544-6115.1309>.
- Lauria, D.T., Hopkins, O.S., Debomy, S., 2005. Pro-poor subsidies for water connections in west Africa. *Water Supply Sanit. Working Notes - World Bank* 45.
- LeDell, E., van der Laan, M.J., Peterson, M., 2016. AUC-maximizing ensembles through metalearning. *Int. J. Biostat.* 12, 203–218. <https://doi.org/10.1515/ijb-2015-0035>.
- Mendoza, G.A., Martins, H., 2006. Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. *For. Ecol. Manag.* 230, 1–22. <https://doi.org/10.1016/j.foreco.2006.03.023>.
- Ministry of Gender, 2022. Children and Social Protection. LEAP Programme. URL. <https://leap.mogcsp.gov.gh/about-us/district/>.
- Ministry of Sanitation and Water Resources - Ghana, 2018. Guidelines for Targeting the Poor and Vulnerable for Basic Sanitation Services in Ghana. Ministry of sanitation and water resources - Ghana.
- O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quantity* 41, 673–690. <https://doi.org/10.1007/s11135-006-9018-6>.
- Panda, S., 2015. Political connections and elite capture in a poverty alleviation programme in India. *J. Dev. Stud.* 51, 50–65. <https://doi.org/10.1080/00220388.2014.947281>.
- Perez, A., Ganguli, S., Ermon, S., Azzari, G., Burke, M., Lobell, D., 2019. Semi-Supervised Multitask Learning on Multispectral Satellite Images Using Wasserstein Generative Adversarial Networks (GANs) for Predicting Poverty arXiv:1902.11110 [cs].
- Podgorski, J., Berg, M., 2020. Global threat of arsenic in groundwater. *Science* 368, 845–850. <https://doi.org/10.1126/science.aba1510>.
- PPI, 2019. Ghana 2016 PPI - user guide. IPA (Innovations for Poverty Action).
- Rao, A., Monteiro, J.M., Mourao-Miranda, J., 2017. Predictive modelling using neuroimaging data in the presence of confounds. *Neuroimage* 150, 23–49. <https://doi.org/10.1016/j.neuroimage.2017.01.066>.
- Remeseiro, B., Bolon-Canedo, V., 2019. A review of feature selection methods in medical applications. *Comput. Biol. Med.* 112, 103375 <https://doi.org/10.1016/j.combiomed.2019.103375>.
- Rodriguez-Galiano, V.F., Luque-Espinar, J.A., Chica-Olmo, M., Mendes, M.P., 2018. Feature selection approaches for predictive modelling of groundwater nitrate pollution: an evaluation of filters, embedded and wrapper methods. *Sci. Total Environ.* 624, 661–672. <https://doi.org/10.1016/j.scitotenv.2017.12.152>.
- Rutstein, S., 2008. The DHS Wealth Index: Approaches for Rural and Urban Areas, DHS Working Papers No. 60. Macro International, Calverton, Maryland, USA.
- Rutstein, S., Johnson, K., 2004. DHS Comparative Reports. The DHS Wealth Index. ORC Macro, Calverton.
- Sasman, C., 2010. Namibia: a case for rural water subsidies. URL. <https://reliefweb.int/report/namibia/namibia-case-rural-water-subsidies>.
- Schnitzer, P., Stoeffler, Q., 2022. Targeting for social safety nets: evidence from nine programs in the sahel. *SSRN* 64. <https://doi.org/10.2139/ssrn.4017172>.
- Stoeffler, Q., Mills, B., del Ninno, C., 2016. Reaching the poor: cash transfer program targeting in Cameroon. *World Dev.* 83, 244–263. <https://doi.org/10.1016/j.worlddev.2016.01.012>.
- Taylor, R., 1990. Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Sonogr.* 6, 35–39. <https://doi.org/10.1177/875647939000600106>.
- Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2014. Predicting water quality at Santa Monica Beach: evaluation of five different models for public notification of unsafe swimming conditions. *Water Res.* 67, 105–117. <https://doi.org/10.1016/j.watres.2014.09.001>.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B* 58, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- Trémolet, S., Kolsky, P., Perez, E., 2010. Financing On-Site Sanitation for the Poor: a Six-Country Comparative Review and Analysis. Water and Sanitation Program technical paper.
- Van Campenhout, B.F.H., 2007. Locally adapted poverty indicators derived from participatory wealth rankings: a case of four villages in rural Tanzania. *J. Afr. Econ.* 16, 406–438. <https://doi.org/10.1093/jae/ejl041>.
- WASHPaLS, 2021. Identifying Households Eligible for a Targeted Sanitation Subsidy in Rural Ghana (Implementation Brief).
- WASHPaLS, 2020a. Assessing the Effects of COVID-19 on Access to Water, Sanitation, and Hygiene in USAID High Priority and Strategy-Aligned Countries: Synthesis Report (Synthesis Report). United States Agency for International Development (USAID). Water, Sanitation, and Hygiene Partnerships and Learning for Sustainability (WASHPaLS).
- WASHPaLS, 2020b. Implementation of a Targeted Toilet Subsidy in Ghana: Midline Report 60.
- Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nat. Commun.* 11, 1–11. <https://doi.org/10.1038/s41467-020-16185-w>.
- Zhao, X., Yu, B., Liu, Y., Chen, Z., Li, Q., Wang, C., Wu, J., 2019. Estimation of poverty using random forest regression with multi-source data: a case study in Bangladesh. *Rem. Sens.* 11, 375. <https://doi.org/10.3390/rs11040375>.