

The (lack of) Distortionary Effects of Proxy-Means Tests: Results from a Nationwide Experiment in Indonesia

Abhijit Banerjee
MIT

Rema Hanna
Harvard Kennedy School

Benjamin A. Olken
MIT

Sudarno Sumarto
TNP2K and SMERU

September 2019

Abstract

Many developing country governments determine eligibility for anti-poverty programs using censuses of household assets. Does this distort subsequent reporting of, or actual purchases of, those assets? We ran a nationwide experiment in Indonesia where, in randomly selected provinces, the government added questions on flat-screen televisions and cell-phone SIM cards to the targeting census administered to 25 million households. In a separate survey six months later, households in treated provinces report fewer televisions, though the effect dissipates thereafter. We find no change in actual television sales, or actual SIM card ownership, suggesting that consumption distortions are likely to be small.

We gratefully acknowledge Bambang Widiyanto, Elan Satriawan, Suahasil Nazara and Priadi Asmanto at the Indonesian National Team for the Acceleration of Poverty Reduction (TNP2K), the Indonesian Central Bureau of Statistics (in particular, Wynandin Imawan and Thoman Pardosi) and the Ministry of Communications and Information (KeMenKomInfo) for their support implementing this project, and thank Talitha Chairunissa, Masyhur Hilmy, Garima Sharma, Sam Solomon, Suhas Vijaykumar, and Poppy Widyasari for research assistance. This project was approved by the MIT IRB (Protocol #1504007043). This research was supported by a grant from the Australian Department of Foreign Affairs and Trade. The views expressed in this paper are those of the authors alone, and do not necessarily reflect the views of any of the institutions or individuals acknowledged here. Contact information: rema_hanna@hks.harvard.edu.

I. Introduction

The past few decades have seen a dramatic expansion in government run anti-poverty programs in developing countries, so that today at least 67 low- and middle-income countries have programs targeted to poor households (World Bank 2015). A key challenge for these programs, however, is identifying which households are poor and hence eligible for the programs. Unlike in the developed world, where governments can target based on incomes reported through the tax system, in developing countries, with large informal sectors, the vast majority of households pay no tax. This means that governments have to come up with other means of identifying the poor.

Instead, in a process called “proxy-means testing,” developing country governments often use infrequently conducted censuses of assets and demographics, predict incomes based on these assets, and then create beneficiary lists based on these predicted incomes.¹ This is a substantial undertaking: every few years, the government sends thousands of enumerators door-to-door throughout the entire country, collecting data firsthand on the assets of in many cases millions of households. While households can infer from the questionnaire which assets the government is asking about, and therefore which assets may affect targeting decisions, the precise formula used to map from the list assets to eligibility is almost always kept secret. This type of proxy-means testing is quite common, used in both large countries such as Pakistan, Nigeria, Mexico, Indonesia, and the Philippines, and in smaller countries, ranging from Burkina Faso to Ecuador to Jamaica (Fiszbein and Schady 2009).

An important policy concern with this approach is that, by conditioning benefits on which assets households own as measured by the proxy-means census, these tests implicitly place a tax

¹ Other methods used for targeting in developing countries include community based targeting (see, for example, Alderman 2002; Galasso and Ravallion 2005; Dupas, Preciado, and Robinson 2018; and Alatas et al. 2012) and ordeal mechanisms (Dupas et al. 2016; Alatas et al. 2016). See Coady, Grosh, and Hoddinott (2004) for an overview.

on ownership of these assets. Prior work has shown that households may strategically misreport on the targeting census itself (see, e.g. Martinelli and Parker 2009 and Camacho and Conover 2011). For example, anecdotally, one often hears that households hide their TV and motorcycles under a tarp when they see surveyors approaching to conduct the targeting censuses. However, to the best of our knowledge, a critical and as yet unanswered question is whether this type of asset-based targeting tax spills over beyond the targeting census itself, and in particular, whether it distorts actual consumption behavior. Such real distortions are particularly important in thinking about the design of anti-poverty programs since they could have real economic effects if the assets themselves are productive (e.g. livestock, cell phones) or if they have potential health effects (e.g. better toilets for sanitation).

In the extreme, these types of distortions could certainly occur. For example, in the 18th and 19th century, England and Scotland taxed windows as an easily observable proxy for the wealth of households, and this famously led to windows being boarded up and countless dark houses (Oates and Schwab 2015). However, modern proxy-means tests use a large number of assets – 34 different types of housing characteristics and assets in our Indonesian example – and as described above governments deliberately keep the formulas that relate assets to eligibility opaque to try to prevent manipulation around their results. On net, the implied marginal tax rate on consumption from Indonesia’s proxy-means tested benefits is about 15 percent (Hanna and Olken 2018), which certainly seems large enough that it could cause distortions in aggregate.² However, the fact that the relationship between assets and eligibility for benefits is complex and non-transparent may mean that the distortionary effects could be small in practice (Chetty et al. 2009; Finkelstein 2009). Indeed, some have argued that on balance the distortions from targeted programs may be small in

² A 15 percent tax is quite large and has the potential to distort behavior. For example, Chetty et al. (2009) showed that consumers responded in magnitudes predicted by theory to a sales tax of 7.375 percent when the tax was posted.

developing country context (Ravallion 2003). On balance, whether these censuses cause actual distortions in asset ownership is ultimately an empirical question.

To answer this question, we conducted a unique nationwide randomized experiment that tested whether Indonesia's *real* targeting census has an impact on subsequent asset acquisition. Indonesia's targeting census occurs approximately every 3-4 years, with enumerators going door-to-door to interview 25 million households—generating data on 92 million individuals—to determine citizens' eligibility for transfer programs and subsidies.

Specifically, we randomized two additional new questions onto the 2015 version of questionnaire, launched in June of that year. To keep the number of questions on the census constant, each randomized question had one of two options. In half the provinces, households received (1) either a question on flat-screen television ownership or a question on the number of rooms in their house and (2) either a question on the number of active cell-phone SIM card numbers the household had or whether they had a modern toilet installed. We specifically chose our two key treatment questions – flat-screen televisions and SIM card ownership –because we had access to independent data sources on actual asset ownership that did not rely on household self-reports. Questions (1) and (2) were cross-randomized to create 4 versions of the census, randomized across Indonesia's 34 provinces to create a nationwide experiment.

The randomization was done at such a large scale – nationwide in scope, and with randomization province by province – in order for the experiment to be as natural possible (Muralidharan and Niehaus 2017). In particular, the Indonesian statistics bureau that administers the targeting census is organized provincially, and hence each province's statistics office, field staff, and data entry team worked with the same version of the targeting census. Moreover, since the government rarely removes questions from one targeting census to the next – for example, 26

of the 29 asset questions (i.e., 90 percent) from the 2011 targeting census were indeed asked again on the 2015 targeting census, and the three that were not included directly were asked again in a slightly different form³ – households could reasonably forecast that these new questions would continue to be asked in future targeting censuses as well.

We test whether these questions led to differences in both *reported* asset ownership, as measured by subsequent government household sample surveys that have no link to targeting, and in *actual* asset ownership, as measured by independent data on television sales that we obtained from retailers and from administrative data on the number of SIM cards active in each province that we obtained from the telecommunications providers.

To test the effect on reported ownership, we worked with the National Statistics Agency (BPS) to include all four of these questions on the Indonesian National Socioeconomic Survey (“SUSENAS”), administered annually to over 250,000 households. (For clarity, we hereafter refer to the targeting census of the poor – which is our treatment – as the ‘census,’ and the subsequent SUSENAS survey data that we use for outcomes as the ‘survey.’) Even though the SUSENAS is not used for targeting—and in fact, the government agencies that conduct targeting cannot even access identifiable data from it—it is possible that individuals may still inaccurately perceive that the government may use it, and thus try to misreport and/or hide their assets. The fact that targeting based on assets may cause misreporting more generally in government data is a substantial concern in itself as this data is used for calculating descriptive measures of the economy, including the local and national poverty rates.

³ For example, one of the three questions not asked verbatim was as follows: the 2011 targeting census asked about 12kg or above LPG (cooking gas) cylinders, whereas the 2015 targeting census asked about 5.5kg or above LPG cylinders.

Analyzing the national sample survey data about six months after the targeting was complete (March 2016), we find that households who live in the provinces where the targeting census asked about flat-screen televisions were in fact 16 percent less likely to report owning a flat-screen television in the SUSENAS. We do not observe any effect of the other questions (toilets, rooms or SIM cards) on reported ownership, but the effect on flat-screen televisions nevertheless survives multiple-inference adjustment. One year later (March 2017), we no longer observe any differences in flat-screen TV ownership across the experimental groups, nor do we observe differences across the other asset variables.

To test whether the targeting questionnaire has distortionary effects on actual asset ownership, we obtained data on television sales from a monthly retailer survey conducted by a leading Indonesian marketing firm, and administrative data on yearly SIM card subscribers from all major Indonesian telecommunications companies through the Ministry of Communications and Information (KeMenKomInfo). We find no evidence of lower television sales, or fewer SIM cards owned, in the provinces in which these questions were asked on the proxy-means tests. Moreover, we can strongly reject a decline in *actual* television sales that would be required to produce the 16 percent decline in reported television ownership detected in the March 2016 SUSENAS; indeed, our estimates suggest that at least 94 percent of the decline we observe in the 2016 SUSENAS is due to reporting, not actual changes. We find no detectable changes in cell phone ownership, though the confidence bands are somewhat larger. The results suggest, therefore, that observed differences in the survey data based on the experimental groups are largely due to effects on reporting, rather than real distortionary effects on asset purchases.

These findings contribute to the policy debate on whether targeted transfers are an effective tool in the fight against poverty. One common argument brought about by critics of targeted

programs is that targeting distorts real behavior and choices, and this in turn could have implications for poverty and growth. Providing experimental evidence embedded into Indonesia's real targeting system, this paper suggests that while there may be strategic responses on subsequent surveys, potentially affecting targeting decisions, the real consumption distortions from avoiding assets used in the targeting formula are indeed likely to be small.

Section II provides the setting, experimental design, and data. We discuss our findings in Section III. Section IV concludes.

II. Setting, Experimental Design and Data

A. Setting

In developed countries, the selection of the beneficiaries for social protection programs (“targeting”) is often accomplished through means-testing: only those with incomes below a certain threshold are eligible. However, for many lower income countries, it is challenging to conduct conventional means-testing as many people work in agriculture or in the informal sector, and thus lack verifiable records of their income.

Instead, to determine program eligibility, many governments conduct “proxy” means testing, where they use demographic and asset ownership data to predict poverty status. Typically, they conduct periodic quasi-censuses of the poor where government enumerators go door-to-door—visiting millions of households—to acquire information about pre-existing household demographic composition and assets, such as the type of material used in the roof or the walls, whether a household owns a refrigerator or a motorcycle, and so on. The government then takes these variables and uses them to predict incomes, usually based on a formula derived from a

prediction exercise using survey data. Program eligibility is then determined by predicted income or per capita expenditures.

Indonesia is no exception. The government has conducted nationwide targeting censuses of the poor approximately every three years since 2005, and has then used proxy-means testing to determine each household's eligibility for targeted transfer programs ranging from cash transfers to health insurance for the poor.⁴ The government canvassed 25 million households, generating data for about 92 million individuals, in the most recent national targeting census—called the *Pemutakhiran Basis Data Terpadu*, or PBDT—in June through August 2015 (see Appendix Figure 1). The three-page targeting questionnaire consisted of three sections: one on basic housing characteristics (e.g. type of roof material, type of floor material, etc.), one on demographics, and one on the assets owned by the household, including items such as refrigerators, A/C, motorbikes, land, and livestock.

The government ran socialization meetings in each village and urban neighborhood prior to conducting the targeting census in which the link between the targeting census and subsequent receipt of government programs was explained. The primary reason for these meetings is that the government wanted to solicit local input to make sure they canvassed all potentially poor households. The briefing materials described the targeting purpose of the survey, but did not contain any details on which questions would be used for targeting or the precise formula.⁵

⁴ To derive the enumeration frame for the census of the poor, Indonesia, like other countries, uses a combination of methods (e.g. past PMT score, community targeting) to exclude rich households from the data collection. Thus, the census of the poor, in practice, covers 25 million households, or about 40 percent of the population.

⁵ For example, a village newspaper from Central Java reporting on the socialization meeting explained that the targeting census “is a collection of direct data to determine beneficiaries of social programs.” Beyond that, they explained that the statistics agency is “only doing the data collection, and the determination of who will receive benefits will be made by the relevant government agencies” and did not provide any additional details on which variables would be used. See, for example, <https://www.wlaharwetan.desa.id/desa-wlahar-wetan-bersama-bps-gelar-forum-konsultasi-publik-pemutakhiran-data-terpadu/>. This echoes the messages in the central government's brochure on the survey: (<http://www.tnp2k.go.id/images/uploads/downloads/leaflet-pbdt-alternatif-tiga%20reduced.pdf>).

B. Experimental Design

We worked with the Government of Indonesia to test whether adding additional questions onto the actual 2015 PBDT questionnaire would incentivize households to reduce asset acquisition in order to maintain their eligibility in the future. We randomized two additional questions onto the actual PBDT questionnaire, reaching 92 million individuals, so that while everyone canvassed received the same number of questions, they were randomly asked different asset questions depending on which province they lived in.

Specifically, each household received (1) either a question on flat-screen television ownership or a question on the number of rooms in their house and (2) a question about the number of active SIM cards the household owns or whether they had a ‘swan neck’ toilet installed (see Appendix Table 1 for complete questions).⁶ These questions were added to the forms at BPS Jakarta, and were treated no differently by the regional offices administering the PBDT targeting census from any other questions. We verified in-person that the forms used in the field followed the randomization protocol in a number of selected provinces.

There are at least two reasons to think that adding questions to the census could distort real behavior. First, as described above, the questions on the poverty census generally do not change much from wave to wave, so a reasonable way to forecast of what will be asked on the next poverty census is through the questions on the current census. Second, households may also be concerned that the government may verify their assets if eligible for the program. For either reason, households may reduce their consumption of these assets following the addition of these questions to the targeting census.

⁶ A ‘swan neck’ toilet is the common Indonesian term for any toilet with a modern plumbing trap (typically known as a P-trap) installed to prevent the venting of sewer gasses back into the house.

It is important to clarify that there is no mechanical effect on our results through differential selection of beneficiaries. This is for two reasons. First, the additional questions were not actually used by the final PMT formula to select beneficiaries for government programs. The fact that these questions were not used was not public; the formula is kept secret and is known only to a few select staff members in Jakarta. Indeed, we confirmed this with an extensive media search, which indeed reveals no mentions of the formula or what variables were or not included anywhere in the Indonesian press. Second, in any case, this could not have *been* publicly known at the time of the 2016 household survey that we use to measure outcomes, because the formula was still being refined internally in 2016 and the final list of beneficiaries was not used until early 2017. However, as described above, it was widely known that the PBDT targeting census, in general, would be used for determining program eligibility (and indeed the government held tens of thousands of meetings, in every village and urban neighborhood of Indonesia, to explain this prior to the survey), and hence a reasonable presumption for a normal household is that all questions in it, including the randomly added questions, would have been used.

The randomization was conducted across the 34 provinces, since the enumerator training and forms used occurred at the province-level (Figure 1). We stratified by 5 regions corresponding to the main Indonesian island groups for additional statistical precision.⁷ The fact that the experiment spans all of Indonesia increases external validity, overcoming the fact that there are significant differences in culture and institutions across Indonesia (Dearden and Ravallion 1988).

C. Data

⁷ The strata are Java, Kalimantan, Sulawesi, Sumatra, and all other provinces.

We use three main datasets for this paper. First, we obtained household-level data from the Indonesian National Socioeconomic Survey (SUSENAS), a semi-annual national survey, representative of the population at the district-times-urban/rural level, conducted by the Government of Indonesia. We use SUSENAS data from after the PBDT targeting occurred—specifically, March 2016 and March 2017, comprising about 300,000 households in each round—to measure whether households report owning fewer of a particular asset if they were asked questions on ownership of that asset on the PBDT (see timeline in Appendix Figure 1). Importantly, not all of our outcomes of interest were initially included on the SUSENAS prior to our study, and thus, we worked with Statistics Indonesia (BPS) to make sure all four questions were included. We also obtained earlier years of the SUSENAS data—from 2005 to 2015—in order to include baseline control variables at the district-times-urban/rural level to gain additional statistical precision. For our purposes, we treat the SUSENAS as repeated cross-sections.⁸

The SUSENAS is a sample survey, where households are interviewed to collect their information. If there is an effect of the treatment on asset acquisition using this data, it could be due to two factors. First, treatment households could actually reduce their asset acquisition or choose not to invest in new assets. Second, their asset ownership may not actually change, but they may lie about it to the surveyors (i.e. “hide their income”). In fact, this is a common concern that one often hears about during the targeting census—people hiding their televisions or motorcycles under a cloth when an enumerator is arriving.

Therefore, for two of our questions, we chose variables that we would be able to verify using independently-sourced data that does not rely on household reports. This allows us to shut off the “lying channel” and only measure real effects on asset acquisition.

⁸ A small number of SUSENAS households constitute a panel in some years, but this is not useful in the study period.

First, we obtained data on monthly television sales of flat-screen televisions—from January 2013 through December 2016—from an Indonesian market research firm. The data captures all flat-screen televisions with screens 30 inches or larger, thus matching exactly the question we added to the targeting census questionnaire. The market research firm collects monthly TV sales data directly from their network of retailers in 20 regions in Indonesia, and accounts for between 85 to 90 percent of total sales of flat-panel TVs 30 inches and above. Given contractual restrictions between the market research firm and the retail firms that supply them data, we were not obtain province-by-province data; instead, the firm was able to provided us with monthly data on total sales in each of our four randomized groups of provinces (i.e. TV-phone, TV-toilet, room-phone, room-toilet).⁹

Second, we obtained data on yearly active SIM cards, by province, for 2015, 2016, and 2017, from the Indonesian Government Ministry of Information and Communications (KeMenKomInfo), who compiled it from administrative data supplied by each of Indonesia’s telephone providers.¹⁰

D. Randomization Check

We report a balance check using data from the March 2014 SUSENAS, i.e., data from the year before the intervention. We focus on demographics (e.g. urbanization status, household size) and variables that are similar to our intervention questions. As shown in Appendix Table 2, out of the

⁹ Broadly speaking, the SUSENAS and retail sales estimates provide similar magnitudes of televisions owned. The SUSENAS estimates that about 11 percent of households own at least one flat-screen television, equivalent to 7.84 million households. Adding up the total television sales from the market research firm from January 2013 to March 2016 yields about 7.4 million TVs sold in that period. These will not be exact since some flat-screen televisions were sold before 2013, they acknowledge that they cover about 85 to 90 percent of the market, some people will own more than one TV, etc. However, the fact that the magnitudes appear broadly similar provides reassurance on the consistency of the datasets.

¹⁰ This includes data from Telkomsel, Sampoerna, 3, and Smartfren. Data from XL and Indostat are for 2017 only.

16 coefficients that we consider, 1 is statistically significant at the 5 percent level, which is consistent with what we would expect based on random chance. Nonetheless, in our regression analysis, we control for baseline characteristics using a double LASSO procedure (Belloni, Chernozhukov, and Hansen 2014) to account for any differences across treatment groups in the sample and to increase statistical power.

III. Results

A. Effects on Self-Reported Asset Acquisition

We begin by examining the impact of receiving the randomized asset questions in the PBDT on each of the four considered assets. Specifically, we estimate:

$$Asset_{hdp} = \beta_0 + \beta_1 TV\ Treat_p + \beta_2 Cell\ Treat_p + \mathbf{X}'_{dp}\boldsymbol{\gamma} + \alpha_r + \varepsilon_{hdp} \quad (1)$$

where $Asset_{hdp}$ is the self-reported asset measure in the post-period for household “ h ” in district “ d ” in province “ p .” We include two dummy variables to indicate which of the randomized questions households received on the targeting questionnaire. Therefore, β_1 provides the causal effect of being randomized to the “TV question” rather than the “rooms” question, while β_2 provides the causal effect of being randomized to the “SIM card” question rather than the “toilet” question. We report standard errors clustered at the province level, our level of randomization. We also report p-values computed using randomization inference at the province level (our unit of randomization), with counterfactuals generated using our original randomization programs, and which are valid even with a finite number of provinces (Young 2019).

While the randomization should ensure balanced groups, one can gain additional statistical precision by including controls of two types. First, we include fixed effects for the 5 regional strata (α_r). Second, we include baseline control variables to account for any differences across treatments groups and improve statistical power. Typically, one would include the baseline data of the

outcome variable for a given observation. However, in this case, we could not do that for two reasons—first, the SUSENAS is a repeated cross-section of households within a given district rather than a panel, and second, we added new questions onto subsequent SUSENAS for the purpose of this study, so pre-period values of these controls are not available. Thus, we instead first coded a total of 1,388 asset variables from the 2007 to 2015 SUSENAS, constructed averages by district and urbanization status (the smallest level at which we can merge this data to the outcome data), and merged these averages into the household survey data. To avoid specification searching, we then selected control variables from this set automatically using the double-LASSO approach of Belloni, Chernozhukov, and Hansen (2014).¹¹

Table 1 provides these results. Panel A does so for our four asset outcome variables from the March 2016 SUSENAS, while Panel B reports results using the March 2017 data.¹² In Columns 1 and 2, we report coefficients on $TV\ Treat_{hdp}$ (β_1) and $Cell\ Treat_{hdp}$ (β_2); in Columns 3 and 4, we report coefficients on the Rooms Treatment ($1-TV\ Treat_{hdp}$) and the WC Treatment ($1-Cell\ Treat_{hdp}$). We hypothesize that each treatment question would have the potential to influence the ownership of the asset in question (i.e. randomized TV question on TV ownership, randomized WC question on WC ownership, etc.), while the non-corresponding questions in each column act as a placebo test (i.e. one would not expect the randomized WC question to necessarily have direct effects on TV ownership). In the next to last row, we report the mean of the dependent variable for

¹¹ The LASSO-selected controls vary from column-to-column, and are listed in Appendix Table 3. Appendix Table 4 replicates Table 1 with the strata, but no baseline controls. The findings are qualitatively similar, but we obtained additional statistical precision with the included controls. In the specification with no controls, the coefficient on televisions in the 2016 remains statistically significant without controls (randomization inference p-value 0.018 without LASSO-selected controls, compared to 0.005 with LASSO-selected controls), but the FWER-adjusted p-value is no longer statistically significant.

¹² In the March 2017, one of our questions (Number of SIM cards) was dropped from the SUSENAS, and instead there is a different question on number of people with active cell phones. We, therefore, use number of people with active cell phones as the outcome in 2017. In Appendix Table 5, we show that using this same question in the 2016 data yields similar conclusions as when one uses the number of SIM cards variable as the outcome in 2016 (Panel A, Column 2 of Table 1).

interpretation, and in the final row, we report p -values adjusted to correct for the family-wise error rate (FWER) to correct for multiple inference across the four key outcome variables (TV variable affects TV ownership; cell variable affects SIM card ownership, etc.) following Westfall and Young (1993) and Anderson (2008). FWER p -values are also calculated using a randomization inference procedure at the province level, so they are correct even in finite samples with 34 provinces.

We first turn to the March 2016 SUSENAS (about six months after the targeting census was completed), the first survey round post-treatment that includes our added questions (Panel A). We find that being randomized to the flat-screen television question on the targeting census leads to a reduction in reported flat-screen TV ownership, but the other randomized questions (WC, rooms, and SIM cards) do not lead to any changes in the ownership of the respective assets. The effect on TV ownership in 2016 is both statistically significant and large in magnitude—being randomized to receive the TV question in the PBDT targeting census leads to about a 16 percent (1.7 percentage point) reduction in reported flat-screen TV ownership in subsequent surveys; this is individually significant at the 1 percent level, and has a FWER multiple-inference adjusted p -value of 0.005.

By March 2017 (Panel B), we no longer observe any significant effects of the experimental treatments on any of the asset questions. In fact, we can easily reject that the magnitude of the treatment effect of the TV treatment in 2017 is the same as in 2016 (p -value=0.005). This implies that any observed effects of the treatments on reported assets may be short-lived.

Why might households lie, and if they choose to lie, why might they choose to lie about owning a flat-screen television? To quantify the gains from lying, we re-estimated the Indonesian government's PMT model, augmented to include each of the 4 new variables, using the 2016

SUSENAS in control areas (i.e. for the television variable, we use provinces which were not asked the television question, and so on).¹³ We then calculate, for each household that actually owns the asset in question, the increase in probability of being below the eligibility cutoff that households would receive by simply changing their response on that one variable.¹⁴ The results, shown in Figure 2, show that lying about television ownership has the highest return for the household, increasing the probability of being eligible for benefits by 12 percentage points for a large number of households; the remaining asset variables would yield between a 3.5 and 7 percentage point increase.

Moreover, the number of rooms and presence of a WC are more easily observed since the government enumerators typically walk from room to room during the surveys, and hence are harder to hide. In contrast, televisions and cell phones are more easily hidden (e.g. hide the TV under a tarp or in a box, keep your cell phone in your pocket). It is also worth noting that cell phones are very common, with households on average reporting about 2.2 SIM cards per household. In contrast, flat-screen televisions are rarer, with only about 11 percent of households owning one, and are more likely to be perceived as distinguishing whether a household is wealthy. While this discussion is of course speculative, it is notable that it is consistent with the patterns we find in the data.

B. Effects on Asset Acquisition Measured in Independent Data

¹³ Specifically, we estimate the PMT regression in 2016 SUSENAS, and calculate the fraction of households eligible in each province / rural-urban unit equal to the percentage of households under 1.5 times the official poverty line. We use the estimated prediction errors in the Indonesian PMT system from Alatas et al. (2016).

¹⁴ For count variables (SIM cards and number of rooms), we calculate the change by reducing one's response by one unit.

The findings from the household survey could be driven by real changes in assets, or they could simply reflect hiding assets from or misreporting assets to the survey enumerators. We therefore turn to the independent data to study real outcomes, thereby shutting off the second channel.

TV Sales Data: Given privacy concerns about releasing province-level data, the firm instead generated for us monthly data on sales by each of our four randomized groups of provinces (i.e. TV-cell, TV-toilet, room-cell, room-toilet). Due to the difference in the data structure (only 4 groups of provinces, but with monthly time series data for each of the 4 groups), we cannot analyze the data using the same specifications as above, but instead estimate:

$$\begin{aligned} \text{LogSales}_{mg} = & \beta_0 + \beta_1 \text{TV Treat} \times \text{POST}_{mg} + \beta_2 \text{Cell Treat} \times \text{POST}_{mg} + \\ & \beta_3 \text{POST}_{mg} + \alpha_g \times m + \varepsilon_{gm} \end{aligned} \quad (2)$$

where LogSales_{mg} is the sales in group “g” at month-year “m,” $\text{TV Treat} \times \text{POST}_{mg}$ and $\text{Cell Treat} \times \text{POST}_{mg}$ are indicator variables that equal 1 if you received that respective question and it is the post-period, and $\alpha_g * m$ is a group-by-month linear trend. The α_g are group dummies, which absorb the main effects of TVTreat_g and CellTreat_g . We use the full data set of data available to us: January 2013 to December 2016.¹⁵

Table 2 provides the results. In Column 1, we estimate Equation (2) using OLS with robust standard errors. However, to account for the time-series structure of the data, we provide two other specifications to deal with potential serial correlation. First, in Column 2, we compute Newey-West (1987) standard errors with 3 lags. Finally, in Column 3 (our preferred specification), we estimate a panel-corrected model with AR(1) disturbances, which accounts for AR(1) serial

¹⁵ In Appendix Table 6, we truncate the data to March 2016 for greater comparability with the time period in Panel A of Table 1 (March 2016 SUSENAS). The findings are unchanged, so we use the full data for the main table.

correlation within panels and allows for correlations in a given month across panels.¹⁶ In the end, all three models produce similar results: we do not observe a reduction in TV sales for those in the treatment group, and if anything, the coefficient is positive.

It is important to note that in Table 2, we are measuring a *flow* (sales of new televisions), whereas in Table 1 we are measuring a *stock* (does the household have a flat-screen television). To compare the two magnitudes, we note that to obtain a 16 percent reduction in the *stock* of televisions reported by households in treatment areas in Table 1 by March 2016, just about 6 months after treatment, the *flow* of television sales in treatment areas would need to have declined by 62 percent, equivalent to a decline in log sales of -0.97. We, therefore, test whether we can reject a decline in log sales of 0.97 (last two rows of Table 2). We can easily rule out declines of that magnitude; indeed we can rule out any decline in log television sales larger in magnitude than about 0.06. In short, the vast majority of the effects – at least 94 percent of the decline – seems to be on *reported* television ownership, not actual sales.

SIM Card Subscriber Data: For SIM card ownership, we have annual data (as of December of each year) for each of the 34 provinces from 2015 to 2017. For each year, we estimate:

$$\text{Log Subscribers}_p = \beta_0 + \beta_1 \text{Cell Treat}_p + \beta_2 \text{TV Treat}_p + \varepsilon_p \quad (3)$$

where Log Subscribers_p is the log of the number of SIM card subscribers in province p , and Cell Treat_p and TV Treat_p are the experimental treatments. Standard errors are robust, and we have one observation per province (the level of randomization) in each panel.

¹⁶ This model, estimated via the *xtpcse* command in Stata, specifies the functional form for the Ω matrix to compute standard errors correctly in the presence of serial correlation within panels and contemporaneous time correlation across panels, and deals with auto-correlation using the Prais-Winston correction. It does not use the Ω matrix fully for estimation in FGLS, and is more conservative than FLGS in small-samples (Beck and Katz 1995).

These results are shown in Table 3.¹⁷ We show results from December 2015 (about 4 months post-treatment) in Panel A, those from 2016 (about 16 months post-treatment) in Panel B, and those from 2017 (about 28 months post-treatment) in Panel C. Column 1 estimates equation (3) above. To obtain greater statistical precision, in Column 2, we add log population in the province as a control, and we additionally add strata fixed effects in Column 3. Across all specifications, the effect of the cell treatment is statistically indistinguishable from zero. Moreover, given that we saw no effects on *reported* cell phone ownership, with much tighter standard errors (we can reject a decline in the number of reported SIM cards of more than 3 percent in 2016 and 2017; see Table 1), a reasonable conclusion is that cell phone SIM ownership did not change either.

IV. Conclusion

While targeted transfer programs have been shown to confer significant benefits, particularly on the health and education of children as well as directly on household consumption, much of the debate surrounding transfers often revolves around whether or not they distort the economic behavior of households. The previous literature has shown that the likely effects of targeted transfers on work behaviors in developing countries are indeed low to non-existent. However, given that much of the targeting is based on asset ownership (rather than income) in these countries, it is of key importance to understand whether these transfers distort household consumption behavior.

Using a nationwide experiment built into Indonesia's real targeting system, covering 92 million individuals, we show that while the targeting may affect short-run reporting of assets, it does not distort real consumption behavior in aggregate. Combined with the growing literature that

¹⁷ In Appendix Table 7, we replicate Table 3 dropping Jakarta because Jakarta SIM cards are easier to obtain elsewhere in the country. The conclusions remain unchanged.

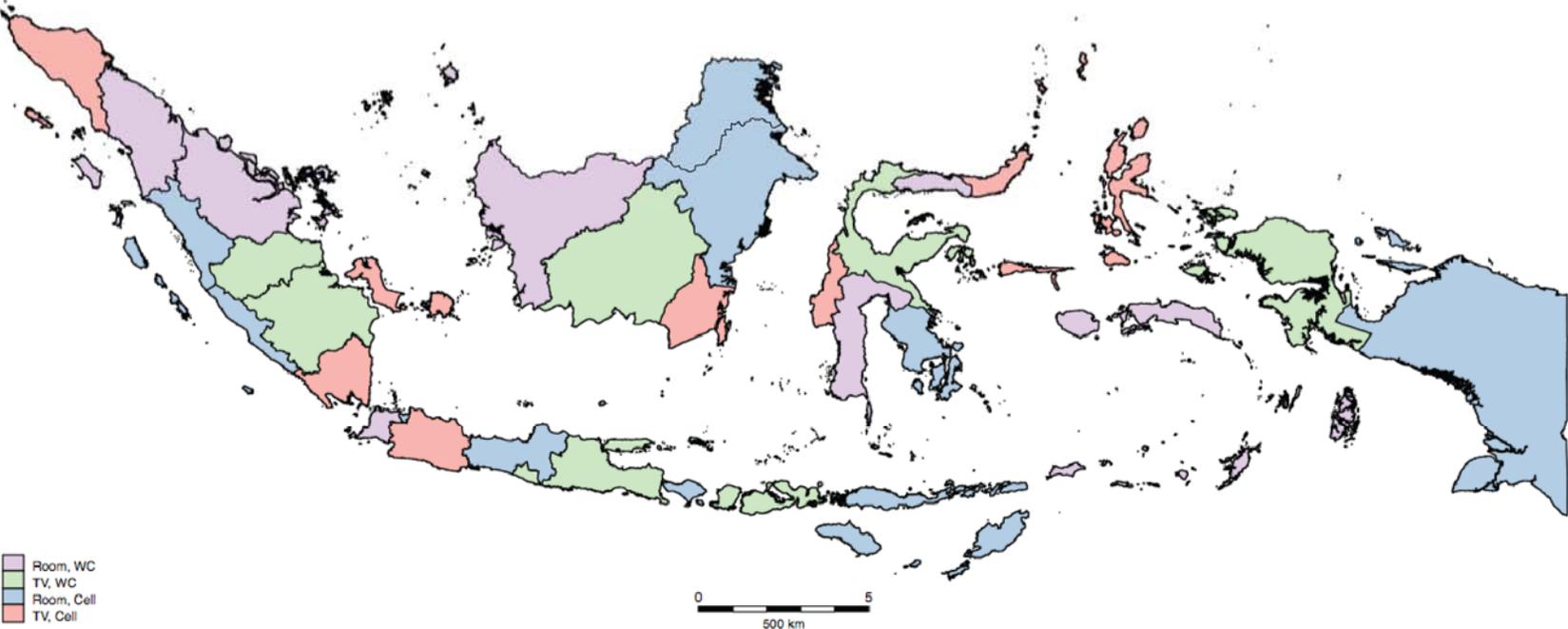
targeting does not greatly reduce labor supply in developing countries, this provides further evidence that the fears around distortive effects of targeting in developing countries is unfounded in the data, especially relative to these potential gains from redistribution.

WORKS CITED

- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2016. "Self-Targeting: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 124 (2): 371–427.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 104 (2): 1206-1240.
- Alderman, Harold. 2002. "Do Local Officials Know Something We Don't? Decentralization of Targeted Transfers in Albania." *Journal of Public Economics* 83 (3): 375-404.
- Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481-495.
- Baird, Sarah Jane, David J. McKenzie, and Berk Özler. 2018. *The Effects of Cash Transfers on Adult Labor Market Outcomes*. Policy Research Working Paper; No. WPS 8404. Washington, D.C.: World Bank Group.
- Banerjee, Abhijit, Rema Hanna, Gabriel Kreindler, and Benjamin A. Olken. 2017. "Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs Worldwide." *World Bank Research Observer* 31 (1): 155-185.
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross Section Data." *The American Political Science Review* 89 (3): 634-47.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29-50.
- Camacho, Adriana, and Emily Conover. 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy* 3 (2): 41-65.
- Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145-77.
- Coady, David, Margaret Grosh, and John Hoddinott. 2004. "Targeting of Transfers in Developing Countries: Review of Lessons and Experience." Washington, D.C.: World Bank.
- Dearden, Lorraine, and Martin Ravallion. 1988. "Social Security in a 'Moral Economy': An Empirical Analysis for Java." *Review of Economics and Statistics* 70 (1): 36-44.

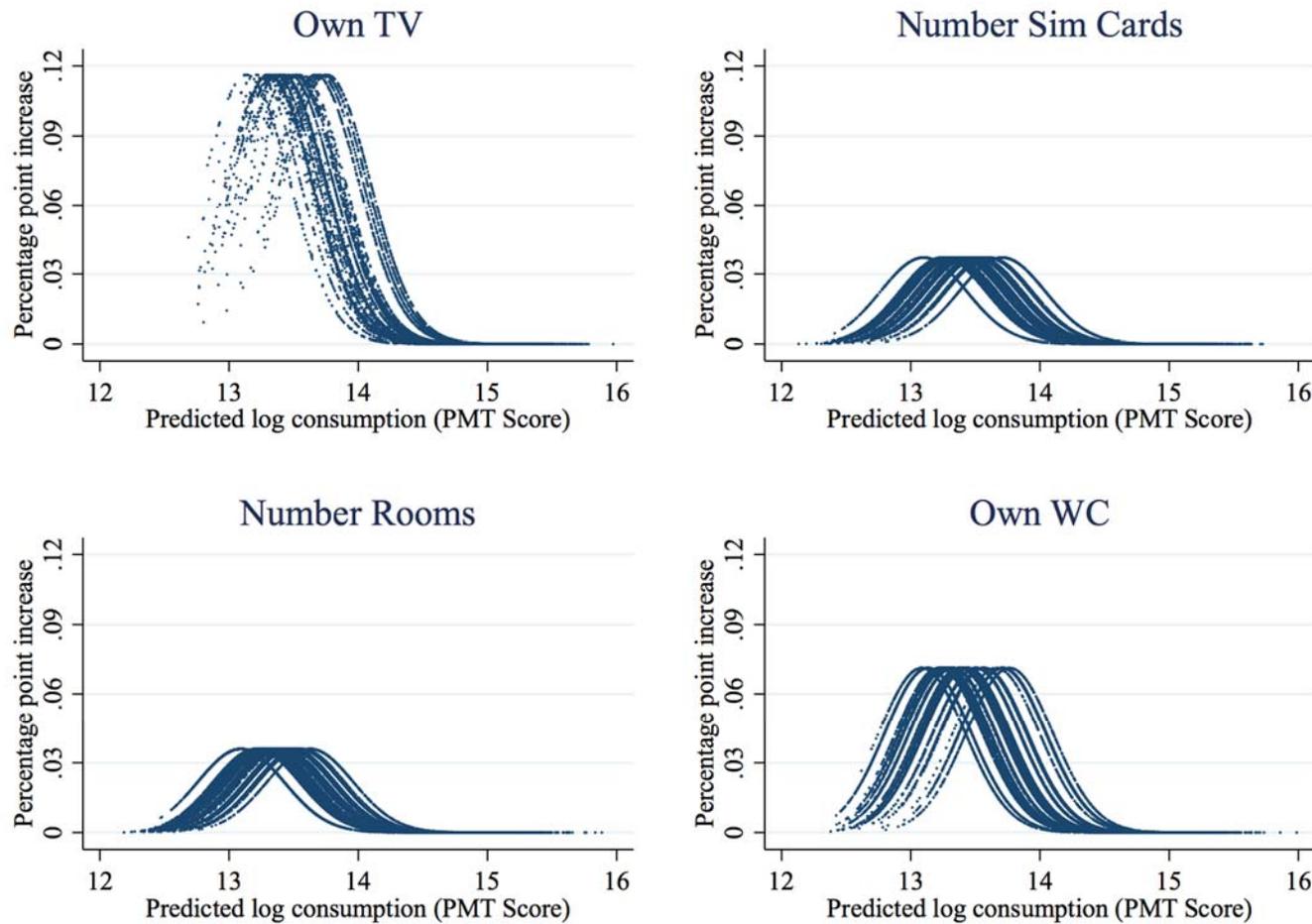
- Dupas, Pascaline, Maria Pia Basurto Preciado, and Jonathan Robinson. 2018. “Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi.” Working Paper.
- Dupas, Pascaline, Vivian Hoffman, Michael Kremer, and Alix Peterson Zwane. 2016. “Targeting Health Subsidies Through a Nonprice Mechanism: A Randomized Controlled Trial in Kenya.” *Science* 353 (6302): 889-895.
- Finkelstein, Amy. 2009. “E-Z Tax: Tax salience and tax rates.” *Quarterly Journal of Economics* 124 (3): 969-1010.
- Fiszbein, Ariel, and Norbert Schady, with Francisco H.G. Ferreira, Margaret Grosh, Niall Keleher, Pedro Olinto, and Emmanuel Skoufias. 2009. “Conditional Cash Transfers: Reducing Present And Future Poverty.” World Bank Policy Research Report.
- Galasso, Emanuela, and Martin Ravallion. 2005. “Decentralized Targeting of an Antipoverty Program.” *Journal of Public Economics* 89 (4): 705-727.
- Hanna, Rema, and Benjamin A. Olken. 2018. “Universal Basic Incomes vs. Targeted Transfers: Anti-Poverty Programs in Developing Countries.” *Journal of Economic Perspectives* 32 (4): 201-226.
- Martinelli, César, and Susan Wendy Parker. 2009. “Deception and Misreporting in a Social Program.” *Journal of the European Economic Association* 7 (4): 886-908.
- Muralidharan, Karthik and Paul Niehaus. 2017. “Experimentation at scale.” *Journal of Economic Perspectives* 31 (4):103-24.
- Newey, Whitney K., and Kenneth D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55 (3): 703-08.
- Oates, Wallace E., and Robert M. Schwab. 2015. “The Window Tax: A Case Study in Excess Burden.” *Journal of Economic Perspectives* 29 (1): 163-80.
- Ravallion, Martin. Targeted transfers in poor countries: revisiting the tradeoffs and policy options. The World Bank, 2003.
- World Bank. 2015. *Closing the Gap: The State of Social Safety Nets 2015*. Washington, DC; World Bank Group.
- Young, Alwyn. 2019. “Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results.” *Quarterly Journal of Economics* 134 (2): 557-598.

Figure 1: Map of Randomization



Notes: This map shows the treatment assignment of each of Indonesia's 34 provinces.

Figure 2: Effect of Misreporting Asset Ownership on Probability of Receiving Benefit



Notes: This figure illustrates households' increase in probability of receiving benefits if they were to misreport asset ownership. Four different Proxy Means Test (PMT) scores are constructed in the control group of each respective outcome, each time using the same categories of variables used in Hanna and Olken (2018), as well as the single relevant asset, as predictors of log per capita consumption. We then calculate the probability of being below the poverty line (defined for each province and for rural and urban areas separately) and thus receiving a benefit, under actual reported assets and under misreporting ownership of the relevant asset. We then graph the difference in these two probabilities as a function of the PMT score, limiting to the households that report owning at least one of the relevant asset so that we do not double count the effect of not owning the asset in the PMT score.

Table 1: Treatment Effect on Self-Reported Asset Acquisition

	(1)	(2)	(3)	(4)
	Own TV	Nb. Sim Cards	Nb. Rooms	Own WC
<i>Panel A: 2016 Outcome Data</i>				
TV Treatment	-0.0171*** (0.00447) [0.005]	-0.00319 (0.0318) [0.932]		
Cell Treatment	-0.00265 (0.00463) [0.727]	0.0125 (0.0277) [0.756]		
Room Treatment			-0.140 (0.179) [0.339]	-0.000209 (0.00468) [0.968]
WC Treatment			0.128 (0.160) [0.372]	0.00587 (0.00468) [0.347]
Observations	291,414	291,414	291,414	291,414
Controls	Lasso	Lasso	Lasso	Lasso
Strata FE	YES	YES	YES	YES
Dep. Variable Mean	0.110	2.183	6.150	0.672
FWER adjusted p-value	0.005	0.735	0.735	0.584
	Own TV	Nb. People with Phones	Nb. Rooms	Own WC
<i>Panel B: 2017 Outcome Data</i>				
TV Treatment	-0.00463 (0.00524) [0.398]	-0.0169 (0.0330) [0.811]		
Cell Treatment	0.00505 (0.00473) [0.371]	-0.0217 (0.0319) [0.725]		
Room Treatment			-0.196 (0.167) [0.238]	-0.00118 (0.00602) [0.862]
WC Treatment			0.0466 (0.152) [0.790]	0.00433 (0.00600) [0.565]
Observations	297,276	297,276	297,276	297,276
Controls	Lasso	Lasso	Lasso	Lasso
Strata FE	YES	YES	YES	YES
Dep. Variable Mean	0.116	1.957	6.229	0.696
FWER adjusted p-value	0.823	0.823	0.722	0.823

Notes: This table provides estimates of the treatment effects of the different targeting questions in the PBDT on household-reported assets in subsequent surveys. Panel A uses data from the March 2016 SUSENAS, while Panel B uses data from the March 2017 SUSENAS. Each regression is estimated using OLS and includes strata from the randomization and pre-experiment control variables selected using the double LASSO procedure, as described in the text. Standard errors are clustered by province; randomization-inference p-values in brackets. The FWER adjusted p-value is calculated within each panel following the free step-down resampling method from Westfall and Young (1993), as described in Anderson (2008) using 1000 bootstraps. *** p<0.01, ** p<0.05, * p<0.1 shown computed using the clustered standard errors.

Table 2: Treatment Effect on Actual Television Sales

	(1) Log Sales	(2) Log Sales	(3) Log Sales
TV Treatment x Post	0.0540 (0.0563)	0.0540 (0.0806)	0.0517 (0.0475)
Cell Treatment x Post	0.190*** (0.0563)	0.190** (0.0806)	0.0771 (0.0505)
Observations	192	192	192
Model/Standard Errors	Robust	Newey	Panel-Corrected AR(1)
Dep. Variable Mean	10.77	10.77	10.77
Treat coeff = -0.97 F-Statistic	331	161.3	463.3
Treat coeff = -0.97 P-Value	0	0	0

Notes: This table provides estimates of the treatment effects of the different targeting questions in the PBDDT on actual television sales. Television sales data is reported monthly from January 2013 to December 2016 for each of the four potential treatment assignments. Column 1 provides simple OLS estimates, while Column 2 provides Newey-West corrected errors with a lag of 3. Column 3 provides panel-corrected estimates with a AR(1) structure. The sales outcome in this table is a flow, while the TV ownership in Table 1 is a stock. To compare the findings, in the final two rows, we also provide a test against the estimate in sales that we would need to observe to generate the ownership effect observed in Table 1. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Treatment Effect on SIM Card Ownership

	(1)	(2)	(3)
	Log Subscribers	Log Subscribers	Log Subscribers
<i>Panel A: 2015 Data</i>			
Cell Treatment	-0.225 (0.406) [0.386]	-0.102 (0.146) [0.643]	-0.106 (0.153) [0.656]
TV Treatment	-0.258 (0.418) [0.323]	-0.192 (0.156) [0.272]	-0.189 (0.148) [0.257]
Observations	34	34	34
Log population control	N	Y	Y
Strata FE	N	N	Y
Dep. Variable Mean	14.95	14.95	14.95
<i>Panel B: 2016 Data</i>			
Cell Treatment	-0.251 (0.401) [0.345]	-0.129 (0.148) [0.556]	-0.135 (0.159) [0.548]
TV Treatment	-0.249 (0.414) [0.356]	-0.184 (0.159) [0.361]	-0.175 (0.152) [0.382]
Observations	34	34	34
Log population control	N	Y	Y
Strata FE	N	N	Y
Dep. Variable Mean	15.17	15.17	15.17
<i>Panel C: 2017 Data</i>			
Cell Treatment	-0.173 (0.403) [0.386]	-0.0478 (0.129) [0.612]	-0.0529 (0.109) [0.570]
TV Treatment	-0.0445 (0.408) [0.817]	0.0227 (0.127) [0.827]	0.0121 (0.103) [0.902]
Observations	34	34	34
Log population control	N	Y	Y
Strata FE	N	N	Y
Dep. Variable Mean	15.55	15.55	15.55

Notes: This table provides estimates of the treatment effects of the different targeting questions in the PBDT on actual active SIM cards subscribers. We have yearly, province level data from 2015 to 2017. All regressions are estimated using OLS, with robust standard errors; randomization-inference p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1 shown computed using the clustered standard errors..