NBER WORKING PAPER SERIES

COMMUNITY TARGETING AT SCALE

Sudarno Sumarto ⓡ
Elan Satriawan ⓡ
Benjamin A. Olken ⓡ
Abhijit Banerjee ⓡ
Achmad Tohari ⓡ
Vivi Alatas ⓡ
Rema Hanna ⓡ

Working Paper 33322
http://www.nber.org/papers/w33322

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

Authors are in random order using the AEA author randomization tool. We thank George Garcia, Karan Makkar, and Eduardo Rivera for outstanding research assistance. Data collection for the data used here was supported by the World Bank, World Bank – Royal Netherlands Embassy Trust Fund, AusAID, and 3ie. We also thank Mitra Samya, the Indonesian Central Bureau of Statistics, the Indonesian Department of Social Affairs, and SurveyMeter for their assistance with the field work for the projects discussed here. This RCT was registered in the American Economic Association Registry for randomized control trials under trial number 99. All views are those of the authors, and do not necessarily reflect the views of any of the individuals or organizations acknowledged here. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33322

Community Targeting at Scale
Sudarno Sumarto ⓡ Elan Satriawan ⓡ Benjamin A. Olken ⓡ Abhijit Banerjee ⓡ
Achmad Tohari ⓡ Vivi Alatas ⓡ Rema Hanna
NBER Working Paper No. 33322
January 2025
JEL No. I38, O15

ABSTRACT

Community-based targeting, in which communities allocate social assistance using local information about who is poor, in experimental settings leads to nuanced allocations that reflect local concepts of poverty. What happens when it is scaled up, by either by making the stakes high, or by replicating the process nationwide? We study this by examining community targeting in both a high-stakes experiment, in which villages determined who would receive the Indonesian conditional cash transfer program – worth almost USD 1,000 over 6 years – and in a nationwide scaleup, whereby Indonesia used community-based meetings to allocate COVID-transfers to over 8 million households. We find that both the experimental scale-up and the massive national scale-up had broadly similar performance to the original experimental study. We find strongly progressive targeting as measured by baseline household consumption, though – as in the pilot – not quite as strong as if they had used a fully up-to-date proxy means test. In both scale-ups, we also find that the villages gave additional weight to locally-valued characteristics beyond pure consumption, such as widowhood, recent illness, and food expenditure shares, again echoing the findings from pilots. The results suggest that community targeting can perform well at scale, as predicted by the experimental study.

Sudarno Sumarto ⓡ
TNP2K
Jl Kebon Sirih no 35
Jakarta 10110, INDONESIA and SMERU
and SMERU
ssumarto@smeru.or.id

Elan Satriawan ⓡ
National Team for Acceleration of
Poverty Reduction (TNP2K)
Grand Kebon Sirih Building, 4th floor
Jl Kebon Sirih no 35
Jakarta 10110, INDONESIA
and Universitas Gadjah Mada
esatriawan@ugm.ac.id

Benjamin A. Olken ⓡ
Department of Economics, E52-542
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
bolken@mit.edu

Abhijit Banerjee ⓡ
Department of Economics,
E52-540 MIT
77 Massachusetts Avenue
Cambridge, MA 02139 and
NBER
banerjee@mit.edu

Achmad Tohari ⓡ
TNP2K
and Airlangga University
ach.tohari@feb.unair.ac.id

Vivi Alatas ⓡ
Asa Kreativita
valatas@worldbank.org

Rema Hanna ⓡ
Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
Rema_Hanna@hks.harvard.edu

A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/99

I. INTRODUCTION

Developing countries face challenges identifying beneficiaries for targeted transfer programs. In high-income countries, governments typically use data on incomes to identify those who are eligible for social assistance. However, in developing countries, high rates of informality and self-employment mean that the government cannot observe income (Jensen, 2023). Instead, many governments identify beneficiaries using a 'proxy-means test' (PMT) – a statistical model that predicts consumption from observable assets. PMTs, while reliable on average, are imperfect: the statistical models does not perfectly predict consumption, generating both inclusion and exclusion errors, where people are falsely included or excluded from beneficiary lists; the underlying asset data is updated infrequently, so the models may not capture recent economic shocks; the models are trained on per-capita consumption, so may miss variation in local concepts of poverty; some households may be omitted from data collection; and the models are complex (and often secret), generating a lack of transparency (Hanna and Olken, 2018).

A different approach, known as 'community targeting', is to devolve the decision about who should receive benefits to local communities. Individuals may have substantial information about their neighbors' incomes, even if this is not observable to a far-off central government. By delegating the beneficiary selection to communities, typically through neighborhood meetings, one can potentially use this local information to overcome some of the challenges of the PMT.

There is, of course, no guarantee this would work – the rich or politically connected could capture the process, for example; people may choose friends, regardless of their income levels; or perhaps people do not actually know their neighbor's income. However, in a small-stakes pilot (Alatas et al, 2012), we found that the community treatment was progressive – and although community targeting performed slightly worse than the PMT in terms of targeting errors with respect to household consumption, it better matched communities' perceptions of poverty and generated higher satisfaction. Community targeting also better identified the poorest relative to the PMT.

What would happen if the government tried to implement community targeting in real-world settings? Would it still lead to progressive targeting when the financial stakes were higher than in the experimental pilot, or the process was rolled out widely across an entire country?

In this paper, we examine the scale-up of different types of "community-targeting" for cash transfers in Indonesia. First, we consider a **high-stakes, experimental** scale-up: Specifically, the government randomized villages to choose beneficiaries of Indonesia's conditional cash transfer program, known as *Program Keluarga Harapan* (Family Hope Program; *PKH*), through either a facilitated community-based process or using a PMT.[1] This program was aimed at the bottom 6 percent of the population and had very high stakes: beneficiaries receive quarterly transfers worth almost USD 1,000 in total spread over 6 years.

Second, we consider a **nationwide scaleup** for cash transfers. During the COVID-19 crisis, the Indonesian government (GoI) was concerned about vulnerable households that may have been excluded from the government's usual targeting procedures for anti-poverty programs, and also for those households whose economic status had deteriorated since the last time that full-scale targeting occurred (2015). Therefore, in April 2020, the government created a new program, in which village governments could use 25-30 percent of the national government's community block grants program to villages (*Dana Desa*) to fund targeted cash transfers. These new transfers were allocated through community-based targeting meetings run by local village governments, in order to identify those who needed help but were not already receiving government assistance. This new program was called *Bantuan Langsung Tunai - Dana Desa* (direct cash assistance – village fund; BLT-DD). By November 2020, over 8 million households had received these community-targeted transfers. Each household received about IDR 2.27 million (~USD$150), and in total, over IDR 18.1 trillion (USD$1.2 billion) was disbursed through these community-targeted funds (Republik Indonesia, 2021).

We first find that both scale-ups – the experimental and the national scale-up – have strongly progressive transfers. In the original pilot study, doubling consumption reduces the odds of receiving benefits by 59 percent. Likewise, in the experimental scaleup, which aimed at the bottom 5-10 percent, doubling income reduces the odds by 68 percent; in the national scaleup, which was aimed at about 18 percent of the rural population (and excluded those already receiving social assistance), doubling income reduces the odds by 38 percent. While both scaleups were progressive, one reason the national scaleup may be somewhat less so is that the scaleup excluded those on other government programs, and hence the very poorest were not eligible. In fact, in the national scaleup, if we look at receiving any government assistance in rural areas – the union of all existing programs and the BLT-DD – doubling income reduces the odds of receiving any benefits by 56 percent, almost identical to our original pilot study.

---

[1] This study also included an on-demand application arm (see Alatas et al, 2016).

Second, we find that both scaleups appear to place more weight on vulnerable households beyond that which would be predicted purely by per-capita consumption – similar to our pilot study. In all three settings, even conditional on per capita consumption, communities are more likely to select widows or those that spend a greater budgetary share on food. Importantly, communities do not appear to discriminate against those from other provinces or districts. The deviations from pure consumption targeting therefore seem to reflect, in part, a social welfare weight that puts weight on these other multi-dimensional aspects of poverty.

Third, while we cannot compare the national scale-up to a PMT (since the counterfactual was not measured), we can compare the community targeting intervention to a PMT that we simulate from baseline data. For comparability, we take this approach in all our datasets – the pilot and both scale-ups. As in the experimental pilot, we find that in both scale-ups, the community targeting would do slightly worse than the simulated PMT in terms of predicting household consumption.

In the large-stakes experimental scaleup, we can also compare experimentally whether the worse targeting performance on consumption reflects improved targeting on other dimensions of local welfare.[2] We find that it does: in both the pilot and the high-stakes scaleup, the results of the community targeting process more accurately reflect how individuals assessed each other compared to the PMT. Perhaps due to this, we find high satisfaction with the targeting process in villages randomized to community-based targeting than in those randomized to the PMT, both in the pilot and the high-stakes experiment.

This paper informs two important literatures. First, we contribute to the literature on community-based targeting (e.g. Olken (2005); Alatas et al (2012); Stoeffler et al (2016); Basurto et al (2020); Premand and Schnitzer (2021); Beaman et al (2021); Dupas et al (2022); Trachtman et al (2023)). We show that community-based targeting can be progressive even on a large scale outside the experimental context, with minimal handholding or even monitoring by national government.

Second, and more broadly, we contribute to the debate on the external validity of experiments. Since the American Economic Association Randomized Control Trial (RCT) Registry was launched in 2013, more than 7,400 RCTs have been registered for over 100

[2] Unfortunately, we lack the data to examine this in the national scale-up, so this analysis is limited to the experimental scaleup.

countries.[3] This has raised debate about whether the results of interventions would hold up as the policies are "scaled up" to more people or areas (e.g. Deaton, 2010; Alcott, 2015; Bold, 2018; Vivalt, 2020).[4] We contribute to this literature by evaluating how the results from a pilot experimental study predict what will happen in other settings, whether it be an experimental integration with a high-stakes government program or in a real-world, national scale-up. The fact that the findings are consistent across the three settings provides a promising example of how smaller-scale studies could be used to inform the program scale-up, despite the fact that the real-world implementations tend to differ from cleaner experimental designs.

II. POLICY INTERVENTIONS AND DATA

A. Community Targeting Pilot Experiment

The original experiment ("Pilot") was conducted in 2008-2009 to help inform whether the GoI should introduce more community-based targeting into its targeting system (Alatas et al, 2012). In 400 communities across Indonesia, we compared two main experimental interventions—a PMT and community-based targeting where villages ranked households from poorest to richest.[5] In addition, we cross-randomized several sub-treatments within the community treatment (i.e. whether the full community was invited or only local leaders, the meeting time of day). This pilot was for relatively small stakes – a one-time transfer of IDR 30,000 (USD 3), which is about one days' wages for a typical worker in the area that we studied.

B. High Stakes Experimental Scale-up

The first scale-up we consider is a high-stakes experimental scaleup ("Experimental Scale-up") that was conducted in 2010-2011. This project grew out of conversations with the GoI explicitly wanting to know if the results of the first experiment would apply in a higher-stakes setting, when the targeting determined the transfers from a real government program. Thus, in the expansion of Indonesia's conditional cash transfer program (PKH) to new areas, we randomized whether the targeting was conducted via a PMT (200 villages) versus community-based targeting (100

[3] https://www.povertyactionlab.org/blog/1-26-24/celebrating-decade-aea-rct-registry
[4] It is worth noting that some experiments are already done at-scale with government partners (e.g. Banerjee et al (2023) or Muralidharan et al (2016)).
[5] In another 200 villages, there was also a hybrid treatment, which was a mix of community and PMT.

villages).[6] The community method in this experiment differed in from the pilot: communities were first shown a list of the poorest households in the village as chosen by the PMT for about 75 percent of the villages' allotment of slots. They were then asked to brainstorm other households that should be in the list, and then conduct a ranking exercise where they could both add and remove households from the government's original list. Following PKH guidelines, the targeting was also aimed at a much poorer sub-population – about 5 percent in our study locations – and households needed to include a child or pregnant mother to be considered. We also cross-randomized whether the full community was invited or just local leaders.[7]

This project represented a dramatically larger scale-up than the original pilot in two senses. First, the stakes were much higher than in the pilot: A typical PKH beneficiary received about IDR 1.5 million per year for 6 years, or almost USD 1,000 over the typical 6-years of eligibility for the program. Second, many more people were considered. The original pilot covered about 12,000 households with about 3,900 households selected. By contrast, in the experimental PKH scaleup, over 120,000 households were considered for community targeting, resulting in over 6,300 beneficiaries.

C. National scale-up: BLT-Dana Desa

During the COVID-19 pandemic, the Indonesian government was concerned about how to reach individuals that were not included in the existing national poverty targeting system but nonetheless faced vulnerabilities. These households may have been missed for several reasons. First, in any targeted system, there can be exclusion error, and there were concerns that these uncovered individuals—already living in poverty—were particularly at-risk during the crisis. Second, the previous targeting system may not have been up-to-date given the large shocks in the economy. The previous targeting was based on a PMT that was last conducted systematically in 2015, with some updates based on local input over the years from 2015-2020. But there could exist people

[6] This experiment had two additional treatments. First, we evaluated a treatment where households could self-select to apply (Alatas, et al 2016). Second, we also examined a hybrid community treatment (the "addition treatment") in another 100 villages, where communities could add, but not subtract from the government's list. We show results from the addition treatment in Appendix Table 2; they are broadly similar.

[7] After the intervention, the Ministry of Social Affairs realized that it had additional funds available and increased the number of program beneficiaries to some additional households that were classified as poor in the 2008 poverty census. We do not include these households as "selected" in our analysis given that they were not selected by the interventions.

whose economic status had deteriorated since the last systematic targeting census and data updating and thus needed help.

The GoI needed a fast and administratively efficient way to channel resources to these groups. In April 2020, they launched Bantuan Langsung Tunai Dana Desa (BLT-DD; Direct Cash Assistance of Village Funds). Specifically, the Dana Desa program (launched in 2015) provides block grants to rural village (desa) governments from the national government; each village receives an average of IDR 1 billion (USD 62,000) per year, and these are used for local public goods, e.g. infrastructure.[8] Under the BLT-DD program, village governments were allowed to allocate up to 25-30 percent of the village funds under the Dana Desa Program to targeted cash transfers for households that were not currently on social assistance programs but were in need. The big question was how to target these funds: there was limited "hard" data (and it would have been challenging to go door-to-door quickly to collect more), there was a need to provide transfers fast, and there was also a need to collect "soft" information on those who may be excluded from a data-driven system.

Although the principles were similar, the detailed protocol for community targeting differed somewhat from the experimental pilot. Village leaders were given a protocol to conduct community-based targeting (Bappenas, 2020), and they did it themselves without external facilitation. The guidelines specified that the village head was to appoint a community targeting committee of at least 3 people (more were allowed, as long as it was an odd number). This committee then went neighborhood-to-neighborhood, consulting with local neighborhood heads (*ketua RW* and *ketua RT*), to assemble a list of poor and vulnerable households. Each household not already receiving transfers was then classified into one of four categories: poor and had recently lost their income; poor and had been excluded from previous lists; poor and chronically ill, and poor and otherwise vulnerable. The village head then convened a public meeting along with the elected village parliament (the *BPD*), which reviewed and certified the list. Household chosen through this process could receive a maximum of IDR 600,000 (USD 45) per month for 3 months and then another 300,000 per month for the subsequent three months.

The program was national and enormous in scope: As of November 2020, the government reported that about 8 *million* households across Indonesia received the program, with BLT-DD

[8] About 69 percent of the Dana Desa funds are equally distributed across villages. The remaining funds are allocated through formulas based on performance, size, etc.

having disbursed IDR 18.14 trillion (USD 1.2 billion) in community-targeted funds (Republik Indonesia, 2021). Note that over 37 million households live in these villages and were eligible (i.e. not on other social assistance programs). Comparing the administrative estimates of the number of beneficiaries with population-weighted estimates from those that report receiving the program in the national sample survey in September 2020, we find that the survey results are consistent with the administrative data, suggesting a majority of people got the program (Hanna et al, 2024). By 2021, the program was in place across 64,886 villages (Republik Indonesia, 2022).[9]

C. Data

For the experimental studies (pilot and scaleup), we use survey data that we independently collected right before the targeting was implemented to understand households' baseline characteristics. Note that the consumption module that we used is taken from the SUSENAS, Indonesia's nationally representative household welfare survey, for comparability. We also conducted endline surveys to better understand program satisfaction.[10]

We put together several datasets to evaluate the BLT-DD. First, to study the targeting properties of the BLT-DD, we utilize data from the March 2020 and September 2020 SUSENAS. The survey is usually run twice per year, with a large wave of approximately 300,000 households (about 1:250 of the population) enumerated every March and a smaller wave of approximately 65,000 households enumerated every September. Typically, the survey is a repeated cross-section. However, given the pandemic, the September 2020 survey re-surveyed a random sample of households from the March 2020 survey.

The panel structure of these data allows us to evaluate how the community targeting fared: From the March 2020 survey, we can observe households' socio-economic status—including demographics, per capita consumption, assets, and whether the households were receiving any other programs—from *right* before the transfers were targeted. From the September 2020 survey, we can identify which households have received BLT-DD transfers.

Finally, we also use data from the 2018 village census (PODES) to provide background data on the three contexts.

[9] As of January 2024, the program is still functioning as a tool to reduce exclusion error within the formal targeting system, but at a small scale, with 10 to 25 percent of village funds being allocated to the program.

[10] Further details can be found in Alatas et al (2012; 2014; 2019).

III. COMPARING POLICY SCALE-UP AND EXPERIMENTAL CONTEXTS

In Table 1, we provide descriptive statistics across the pilot, experimental scale-up, and national scaleup to illustrate the differences in scope, scale, and the socio-demographic characteristics across the three contexts.

Panel A of Table 1 examines program characteristics. Three key differences stand out. First, we observe differences in the locational spread and scale (see Appendix Figure 1). The Pilot community targeting intervention was conducted in one hamlet in each of 200 villages spread across 12 districts. About 3,900 households were selected, out of about 12,000 households. In the experimental PKH scaleup, the community targeting intervention was conducted within 100 villages across 6 districts, but within these villages, every hamlet was targeted. Thus, a total of 6,320 households were selected, out of over 120,000 households. In contrast, the BLT-DD occurred in 426 districts, or about 83 percent of Indonesia's districts. More than 8 million households were selected from an eligible population of over 37 million households.

Second, there are differences in who was targeted. In the Pilot, the bottom 30 percent of the population in terms of per capita consumption were targeted. In contrast, only the bottom 6 percent were targeted in the PKH expansion; households also had to have a pregnant woman or children aged 15 or below. In contrast, the national scale-up aimed to reach those who had not been on social assistance but were nonetheless poor or at-risk for poverty (according to our survey data estimates, this covered about 16 percent of the rural population, or about 18 percent not on social assistance).

Finally, there are large differences in the benefit levels. The Pilot provided a one-time transfer of USD 3 (that was set up specifically for the study). The PKH expansion experiment was higher stakes: almost USD 1,000 in total over 6 years. The BLT-DD was meant to provide short-run assistance, providing a maximum of about USD 186 total spread over 6 months (with communities having some discretion of how much to give, up to the maximum).

Given the differences in locational spread, in Panel B, we next compare district characteristics. We use data from the 2018 village census (PODES) and include all villages in these districts. One clear difference is how urban the districts are: in the Pilot and Experimental PKH scale-up, we stratified the districts to be about half urban. In contrast, the national scale-up districts are only about 28 percent urban, because cities are excluded by construction.

Finally, in Panel C, we compare the household characteristics for the eligible populations. For both experiments, we examine comparable baseline variables from those studies. For the national scaleup, we use the March 2020 SUSENAS, dropping ineligible households (i.e. those on other programs). We find some similarities across the contexts: For example, we observe a similar percent of widows and of households with a disabled member across the groups. However, we also observe some key differences: For example, even adjusting for inflation, the BLT-DD-eligible tend to have higher per capita consumption (IDR 985,594) than those in the two experiments (IDR 733,072 and 720,715 respectively for the pilot and PKH).

IV. FINDINGS

A. Targeting Performance

We start by examining the degree to which community targeting was progressive in the scale-ups, and how that compares to the pilot. In Figure 1, we graph the probability that a household received the program by log per capita consumption out of the eligible population using a locally-weighted regression; we also graph the distribution of per capita consumption in the background for context.[11] Figure 1 Panel A plots the pilot, Panel B plots the experimental PKH scaleup, and Panel C plots the national BLT-DD scaleup for those not on other programs. In addition, to consider the total targeting effects of the system as a whole, in Panel D, we consider all rural households and define receiving benefits as receiving either BLT-DD or other government anti-poverty programs.

Across all programs, the graphs are clearly downward sloping, suggesting that the communities are identifying households that are poorer. In fact, the pattern observed in the BLT-DD scale-up looks similar to that in the interventions conducted in the experiments (top two panels), even though they are targeting different parts of the consumption distribution and have different numbers of slots available.

Table 2 more formally compares the community targeting results across the interventions. We estimate using the conditional logit:[12]

$$Y_{iv} = \beta_0 + \beta_1 X_{iv} + \gamma_v + \epsilon_{iv}$$

[11] In Appendix Figure 2, we plot the probability of program receipt by log per capita consumption after it has been residualized by village fixed effects and observe similar results.
[12] Appendix Table 1 replicates Table 2 using OLS and comes to similar conclusions.

where Y_{iv} is an indicator for whether individual i in village v received the program, X_{iv} is either log per capita consumption or whether one is below the specific poverty line that is being targeted in each program, and γ_v is a village fixed effect. Standard errors are clustered by village.

As in the figures, community targeting is progressive. As shown in Panel A of Table 2, those with higher consumption are less likely to receive any of the programs. Although the magnitude of this relationship is qualitatively smaller in the national scale-up (Column 3) than the pilot (Column 1) when we consider the BLT-DD program alone (and exclude those who were already receiving other government programs), it is almost identical to the pilot when we consider the full sample and examine the combined effect of BLT-DD in combination with other government programs (Column 4). The high-stakes experiment also shows even stronger pro-poor targeting (Column 2). To interpret magnitudes, note that a doubling of income would reduce the odds of obtaining benefits by 59 percent in the pilot (which targeted approximately the bottom 30 percent of households), 68 percent in the experimental PKH scaleup (which targeted approximately the bottom 6-10 percent), 38 percent in the national BLT-DD scaleup alone (which targeted approximately the bottom 18 percent other than those who already received other social assistance), and 56 percent for the combined BLT-DD and other government programs. Similarly, in Panel B, we observe that those below the poverty line are more likely to receive each of the three programs.[13]

We next explore heterogeneity in location within the three programs. First, culturally and economically, Java (which comprises about half of Indonesia's population) differs from other islands; in the experiments, we stratified by on and off Java to examine if community targeting works differently across both locations. In Appendix Table 3, we replicate Table 2 disaggregated by Java and non-Java islands, and find targeting to be progressive on and off Java for both the experiments and the BLT-DD. The negative association between consumption and benefit receipt is larger off-Java than on-Java for both the experimental and national scale-ups ($p < 0.05$).

[13] Appendix Table 2 replicates Table 2, but includes: community treatments disaggregated by whether it was the elite subtreatment or the full community and the addition treatment for PKH. First, the BLD-DD is conceptually most similar to the elite subtreatments, where local leaders made decisions on behalf of the community. We find similar effects in both the elite and full community subtreatments in the experiments. Second, the addition treatment finds smaller effects than in the regular PKH community treatment; this makes sense since it was more constrained in terms of the choices that the communities had.

We then explore whether we observe the same effects when communities have an exogenously greater number of slots to distribute out. Note that we can only do this within the national scale-up. Specifically, nearly 70 percent of Dana Desa funds is an equal grant across all villages regardless of size. Therefore, larger villages can give out transfers to a smaller percentage of households than smaller villages, just as in Kaboski and Townsend (2012)'s study of the Million Baht Fund in Thailand.[14] Interestingly, as one can observe in Appendix Figure 3, even though the probability of program receipt are different in levels due to the differences in funding levels, the slopes of the two lines are quite similar, suggesting that the targeting properties of community-based targeting are not strongly related to the *level* of assistance being given out.

B. Local preferences

We next explore *what characteristics* the community targets. These could be very different across the contexts: both the pilot and scaleup experiments were done in several districts that could have had specific preferences on what they wanted to target on, while the national scale-up was done nationally and so it naturally includes a larger range of preferences or beliefs about who is poor. In Table 3, we estimate the following using a conditional logit:[15]

$$Y_{iv} = \beta_0 + \beta_1 PCC_{iv} + \sum_j \beta_j X_{iv} + \gamma_v + \epsilon_{iv}$$

where we regress benefit receipt on per capita consumption adjusted at 2020 prices (PCC_{iv}), a series of baseline household characteristics (X_{iv}), and village fixed effects (γ_v). We include the baseline characteristics from Table 12 of Alatas et al (2012) that are present in all three datasets.

We find some striking similarities on what communities target on. First, communities choose households that spend a greater share of their consumption on food, even conditional on consumption levels; the effects are similar in both the pilot PKH scaleup, and combined national scaleup, and smaller but still positive in the national scale-up when we examine BLT-DD alone. Second, and importantly, households do not appear to significantly discriminate against households from other provinces or districts, in any of the contexts; the combined national scaleup

[14] Kaboski and Townsend use the fact that each village received 1 million Baht to start a rotating credit funds regardless of village size, so on a per-capita basis, larger villages received a smaller credit injection and vice-versa.
[15] Appendix Table 4 replicates Table 3 but with OLS estimation.

does include fewer new households, but this may be due, in part, to the PMT (which had determined the other programs beyond BLT-DD) being conducted in 2015.

Third, communities appear to choose people who may be more vulnerable on metrics beyond consumption. For example, across all, communities are more likely to choose widows. In both the experimental PKH scaleup and the national scale-up, communities are also more likely to choose those who have experienced illness, although we do not observe this in the pilot. Interestingly, in the national scaleup, the BLT-DD communities were more likely to choose households that have a head with a primary education or less (while positive, we cannot distinguish it from zero in either the pilot or PKH scaleup experiments, though the point estimate is similar in the PKH scaleup and the national scale-up).

On the other hand, there are some clear differences in who the communities chose across the experiment and the scale-up, which are consistent with the program goals in each case. For example, for the experimental PKH scaleup, which was a conditional cash transfer focused on kids, households with a greater share of children were more likely to be chosen. In contrast, in the national scale-up, we observe households with more children (conditional on household size) less likely to be chosen; this may be consistent with the idea that children are cheaper than adults in terms of food consumption, and so households with equal consumption but a higher fraction of kids are likely more food secure.

In contrast, in the national scale-up, we find that those with a greater share of the baseline income from remittances were more likely to be chosen by the community; this was not the case in the experiments. One reason may be that the scale-up was designed to address the COVID-19 crisis where households relying on remittances may have been particularly affected by the crisis.

C. Comparing the Community Targeting Intervention to a Proxy Means Test

In both the pilot and experimental PKH scaleup, we randomized villages to target either using community targeting or a PMT, and we can compare outcomes experimentally across these two groups. For the national scaleup, since the government has scaled up community targeting for the BLT-DD everywhere, we do not have a "real" counterfactual of what would have happened if the government had conducted a proxy means test. However, given that we have baseline data on assets, we can simulate a proxy-means test, and see how the differences in the two methods compare to the two experiments.

To simulate the PMT for the national BLT-DD scaleup, we use as a training database the Hanna and Olken (2018). As some of the variables that were included in the PMT formula in that paper are no longer in the SUSENAS in March 2020, we re-run the prediction formulas subject to data availability.[16] We then apply the formula to March 2020 data, and choose those who have the lowest score in each village up to the number of transfers that the village gave out (since the number of transfer recipients was determined by budget). Since the timing of this asset data matches exactly the consumption data, one can view this as an upper bound on what one would achieve with an actual PMT where the timing may not be the same.

For the pilot and experimental PKH scaleup, we have the original PMT decisions from the experiment. It is worth noting that a simulated PMT has less targeting error than the real PMT conducted by the government, since it is easier to maintain a higher data quality in a smaller survey than a nationwide census (Alatas et al, 2019). Thus, for the experiments, we additionally simulate the PMT using the baseline data from each respective experiment for comparability to the BLT-DD analysis.

We present the results in the first 3 columns of Table 4. Panel A presents the results of the pilot, Panel B presents the results of the experimental PHK scaleup, and Panel C presents the results of the national BLT-DD scaleup, and Panel D for the combined national scale-up. Note that for the pilot and PKH experiment, we examine three PMT versions: the experimental PMT (column 1), the PMT simulated from the baseline (column 2), and a PMT simulated from the baseline but also equalizing the number of people chosen in the PMT and the community treatment (column 3). For the national BLT-DD scaleup and the combined national scale-up, we only have the third option (simulated PMT from baseline survey, equal shares treated), so we focus on this. The key result here is that the community targeting method has slightly worse outcomes in terms of consumption in all three settings: those chosen by the community treatment are about 6 percent richer in the pilot compared to PMT; about 10 percent richer in the experimental PKH scaleup, about 13 percent richer in the national BLT-DD scaleup (when we consider BLT-DD in isolation), and about 5 percent richer in the combined national scaleup. Thus, while community targeting is strongly progressive in all settings, it is not quite as progressive as the PMT.

[16] The PMT includes 68 variables. We dropped bicycle ownership and household pays for drinking water, as these variables are not included in March 2020. We additionally included a dummy variable for those who are unemployed.

The results in Table 3 suggest that the somewhat worse consumption targeting performance may be because communities are targeting other dimensions of poverty.[17] A key question is whether the targeting on other dimensions of poverty is quantitatively large enough to drive these differences between the community intervention and the PMT.

In the experimental PKH scaleup, where we collected our own rich data, we can explore this in two ways. First, as the pilot, as part of the baseline survey, we asked each household whom we interviewed to rank every other household in the neighborhood we worked in from least well off to most well off. We then ask: how do the households chosen to receive the PKH program by the targeting process (community meetings vs. PMTs) compare in terms of their peer welfare ranks? The answer, shown in columns 4 – 6, is that households selected by for PKH are ranked as less well off by their peers at baseline in the experimental scaleup, just as they were in the original small-scale pilot. Similar to the pilot results, this suggests that the slightly worse consumption performance does not reflect worse information or elite capture (in which case we would also expect less of a relationship with baseline peer ranks), but rather an affirmative decision to target local perceptions of poverty.

In the remaining columns of Table 4, we also consider whether the community targeting process leads to greater levels of satisfaction in the experimental PKH scaleup (we do not observe these outcomes for the national BLT-DD and combined scaleup), as we found in the original pilot. In the experimental PKH scaleup, we find greater levels of satisfaction with the targeting process in community villages compared to PMT villages, both among households (columns 7 and 8) and from the village head (column 12). These findings are remarkably similar to what we found in the pilot (shown in Panel A), despite the differences between the two. The only difference in satisfaction is that in the pilot, when we showed households the list of targeted beneficiaries, they reported fewer changes they would have liked to make in community targeting compared to PMT (columns 9-11, Panel A); we do not see analogous results in the experimental PKH scaleup (Panel B). But overall – the comparisons in Table 4 show remarkable similarities between the original pilot and the two experimental scaleups.

V. CONCLUSION

[17] Appendix Table 5 shows the same findings with alternative specifications.

Do experimental results "hold up" when taken to scale? This question is complex, particularly because scale-ups typically do not just involve "xeroxing" the program to exactly the same places in the same form. When programs scale, there's often new policy actors and implementors involved, changes to accommodate the scale, differences across locations, etc.

In this paper, we examine how a community targeting intervention worked in the context of a small-scale pilot in Indonesia, a much larger-scale experiment with high stakes, and a national roll-out. Despite the differences in the programs and implementation, we find that the pilot results are broadly consistent with the results seen in the larger scale experiments–providing a promising example of how experimental results can scale. Specifically, we find that community targeting is progressive in all three experiments, and that communities appear to place greater weight in their decisions on those who may be more vulnerable. Comparing community targeting with a simulated PMT, we find that community targeting fares slightly worse than the PMT in choosing the lowest consumption households across the three contexts. However, the results in the high stakes policy experiment confirm that communities are targeting people that are ranked by their peers as less well off, resulting in higher overall satisfaction with the targeting process. This shows the tradeoffs that governments may make in aiming to capture other benefits of community targeting, i.e. the targeting reflecting local preferences and the lower administrative costs.

VI. WORKS CITED

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, 102: 1206–1240.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, Matt Wai-poi, and Ririn Purnamasari. 2014. International Initiative for Impact Evaluation: Impact Evaluation Report 12. https://www.3ieimpact.org/evidence-hub/publications/impact-evaluations/targeting-poor-evidence-field-experiment-indonesia

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2019. "Does elite capture matter? Local elites and targeted welfare programs in Indonesia." Vol. 109, 334–39.

Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna. 2016. "Self-targeting: Evidence from a field experiment in Indonesia." Journal of Political Economy, 124(2): 371–427.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, Parienté William, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." Science, 348: 1260799.

Banerjee, Abhijit, Rema Hanna, Benjamin A Olken, Elan Satriawan, and Sudarno Sumarto. 2023. "Electronic Food Vouchers: Evidence from an At-Scale Experiment in Indonesia." American Economic Review, 113(2): 514–547.

Bappenas. Panduan Pendataan Bantuan Langsung Tunai–Dana Desa (BLT-Dana Desa), June 2020. Available at: https://sikompak.bappenas.go.id/storage/app/uploads/public/624/2b1/59a/6242b159ae56b332087183.pdf#page=40.25.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson. 2020. "Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi." Journal of Public Economics, 185: 1880–1917.

Beaman, Lori, Niall Keleher, Jeremy Magruder, and Carly Trachtman. 2021. "Urban Networks and Targeting: Evidence from Liberia." AEA Papers and Proceedings, 111: 572–576.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, and Justin Sandefur. "Experimental evidence on scaling up education reforms in Kenya." Journal of Public Economics 168 (2018): 1-20.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature*, 48 (2): 424-55.

Dupas, Pascaline, Marcel Fafchamps, and Deivy Houeix. 2022. “Measuring Relative Poverty through Peer Rankings: Evidence from Côte D’Ivoire.” National Bureau of Economic Research.

Hanna, Rema and Olken, Benjamin, 2018. “Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries.” Journal of Economic Perspectives, 32(4), pp.201-226.

Hanna, Rema, Benjamin A Olken, Sudarno Sumarto, Achmad Maulana, Vivi Alatas, and Elan Satriawan. 2024. “The Challenges of Pre-Employment Assistance: Nationwide Evidence from Indonesia.” Manuscript.

Meager, Rachael. 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." American Economic Journal: Applied Economics, 11 (1): 57-91.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. “Building state capacity: Evidence from biometric smartcards in India.” American Economic Review, 106(10): 2895–2929.

Premand, Patrick, and Pascale Schnitzer. 2021. “Efficiency, Legitimacy, and Impacts of Targeting Methods: Evidence from an Experiment in Niger.” The World Bank Economic Review, 35: 892–920.

Republik Indonesia. 2021. “Buku II Nota Keuangan Beserta Anggaran.” https://web.kemenkeu.go.id/media/16834/buku-ii-nota-keuangan-beserta-apbn-ta-2021.pdf

Republik Indonesia. 2022. “Buku II Nota Keuangan Beserta Anggaran Pendapatan Dan Belanja Negara.” https://web.kemenkeu.go.id/media/16834/buku-ii-nota-keuangan-beserta-apbn-ta-2021.pdf

Stoeffler, Quentin, Bradford Mills, and Del Ninno Carlo. 2016. “Reaching the Poor: Cash Transfer Program Targeting in Cameroon.” World Development, 83: 244–263.

Trachtman, Carly, Yudistira Hendra Permana, and Gumilang Aryo Sahadewo. 2023. “How Much do our Neighbors Really Know? The Limits of Community-based Targeting.” Manuscript.

Vivalt, Eva. 2020. “How Much Can We Generalize from Impact Evaluations?” Journal of the European Economic Association, 18(6): 3045-3089.

Table 1: Comparison of Program Characteristics and Locations

	Pilot *(BLT)* (1)	*Experimental Scale-up* *(PKH)* (2)	*National Scale-up* *(BLT-DD)* (3)
Panel A: Program Characteristics			
Cash Transfer (USD)	\$3	\$17-\$62	\$31
Transfer Periodicity	Once	Quarterly	Monthly
Transfer Duration		$\approx$ 6 years	6 months
Transfer Value (USD)	\$3	\$970	\$186
Number of Recipients (HH)	3,887	6,320	8,040,000
HH Considered for Targeting	11,825	122,785	37,076,079
Number of Districts	12	6	426
Panel B: Districts Characteristics			
Urban	0.494	0.483	0.284
Kelurahan	0.380	0.399	0.151
% households in agriculture	0.434	0.423	0.529
Number of secondary schools/1000 people	0.357	0.266	0.460
Number of mosques/1000 people	0.969	1.008	0.888
Number of base transceiver stations/1000 people	0.238	0.217	0.289
Panel C: Household Characteristics			
log PCE, 2020 IRP	13.505	13.488	13.801
Share PCE to food	0.530	0.593	0.566
Below poverty line	0.158	0.133	0.085
Household size	4.389	4.761	3.622
Share children	0.270	0.301	0.226
HH head with primary education or less	0.552	0.621	0.479
HH head from other province	0.051	0.151	0.141
HH head from other district	0.191	0.105	0.087
HH head immigrated <5 years ago	0.091	0.059	0.028
Widowed	0.209	0.182	0.193
Disability	0.065	0.065	0.075
Sick	0.163	0.153	0.226
Has savings account	0.351	0.370	0.582
Remittances main source of income	0.051	0.062	0.058
Observations	1925	966	36395

Notes: This table provides information about the program characteristics and locations across the two experiments (Columns 1 and 2) and the scale-up (Column 3). Panel A provides basic program characteristics, while Panel B compares the characteristics of the districts using the 2018 village census (PODES). Finally, in Panel C, we provide baseline statistics of households in the areas that the programs were administered. For the Pilot and PKH Expansion Experiments, the data are drawn from baseline surveys collected by the researchers. For the BLT-DD, the data are drawn from the March 2020 SUSENAS. We only include households that are within the community's set to choose from: thus, we drop households automatically included in the Addition Treatment of the PKH Extension analysis and we drop households that are beneficiaries of other programs (e.g. PKH, KKS) in BLT-DD analysis.

Table 2: Determinants of Benefit Receipt

	Pilot	*Experimental Scale-up*	*National Scale-up*	
	(BLT)	*(PKH)*	*(BLT-DD Only)*	*(Combined)*
	(1)	(2)	(3)	(4)
Panel A: log PCE				
log PCE, 2020 IRP	-1.287***	-1.662***	-0.694***	-1.175***
	(0.116)	(0.282)	(0.037)	(0.028)
Mean if no receipt	13.640	13.527	13.840	13.823
Observations	1667	655	22069	41932
Panel B: Below poverty line				
Below poverty line	1.145***	1.092***	0.491***	0.835***
	(0.163)	(0.313)	(0.065)	(0.043)
Mean if no receipt	0.106	0.100	0.066	0.069
Observations	1667	655	22069	41932

Notes: This table presents the results of a conditional logit regression of program receipt on log household per capita consumption (Panel A) and whether the household falls below the province-urban/rural poverty line (Panel B), controlling for village fixed effects. Household per capita consumption is in 2020 price levels. See Table 1 for additional notes on the data and the sample restrictions. Column 4 keeps households that are beneficiaries of other programs (e.g. PKH, KKS) in the BLT-DD analysis. Standard errors are clustered at the village level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Relative Importance of Baseline Household Characteristics in Benefit Receipt

	Pilot	*Experimental Scale-up*	*National Scale-up*	
	(BLT)	*(PKH)*	*(BLT-DD Only)*	*(Combined)*
	(1)	(2)	(3)	(4)
log PCE, 2020 IRP	-1.029***	-0.760**	-0.585***	-0.891***
	(0.152)	(0.349)	(0.049)	(0.037)
Share PCE to food	2.120***	2.058*	0.768***	1.859***
	(0.577)	(1.207)	(0.190)	(0.139)
Household size	-0.185***	0.132	-0.040***	0.072***
	(0.047)	(0.092)	(0.015)	(0.011)
Share children	-0.113	3.201***	-0.216**	-0.489***
	(0.366)	(0.852)	(0.098)	(0.070)
HH head with primary education or less	0.046	0.458	0.233***	0.496***
	(0.169)	(0.315)	(0.036)	(0.025)
HH head from other province	-0.229	0.098	-0.038	-0.073
	(0.367)	(0.402)	(0.066)	(0.049)
HH head from other district	0.062	-0.371	-0.069	-0.101**
	(0.190)	(0.355)	(0.070)	(0.051)
HH head immigrated <5 years ago	0.412	-0.651	0.149	-0.257***
	(0.312)	(0.697)	(0.103)	(0.082)
Widowed	0.789***	0.531*	0.290***	0.200***
	(0.182)	(0.293)	(0.043)	(0.031)
Disability	-0.386	0.804*	0.084	0.068
	(0.240)	(0.444)	(0.064)	(0.046)
Sick	0.172	0.557*	0.125***	0.088***
	(0.210)	(0.289)	(0.039)	(0.027)
Has savings account	-0.682***	-0.195	-0.426***	0.285***
	(0.175)	(0.291)	(0.043)	(0.031)
Has motor vehicle	-1.380***	-0.922***	0.003	-0.249***
	(0.157)	(0.287)	(0.051)	(0.036)
Remittances main source of income	-0.343	-0.780	0.151**	0.235***
	(0.328)	(0.607)	(0.074)	(0.053)
Temporarily unemployed	0.391*	0.445	0.130	0.069
	(0.229)	(0.357)	(0.103)	(0.078)
Avg. receipt prob.	0.312	0.148	0.184	0.184
Observations	1667	655	22069	41932

Notes: This table examines the baseline characteristic predictors of whether one receives benefits across the three programs. We estimate a conditional logit of benefits receipt on the given variables and a set of village fixed effects. We included variables (when available across datasets) from Table 12 of Alatas et al (2012). See Table 1 for additional notes on the data and the sample restrictions. Column 4 keeps households that are beneficiaries of other programs (e.g. PKH, KKS) in the BLT-DD analysis. Standard errors are clustered at the village and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

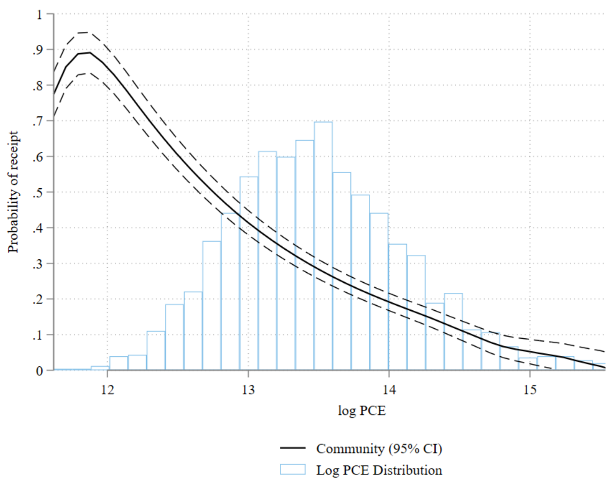
Table 4: Program Recipient Log PCE and Community Rank Percentile by PMT vs Community Targeting

	Targeting Outcomes						Satisfaction Outcomes					
	Consumption			*Community Ranking*			*Targeting method appropriate?*	*Overall satisfaction*	*Any HH missed?*	*Num exclusion error*	*Num inclusion error*	*Vilage head: community satisfaction*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Pilot (BLT)												
Community Targeting	0.031	0.046	0.059**	-0.070***	-0.089***	-0.071***	0.166**	0.245***	-0.191***	-0.578***	-0.559***	0.939***
	(0.0381)	(0.0368)	(0.0183)	(0.0116)	(0.0116)	(0.00791)	(0.0621)	(0.0593)	(0.0393)	(0.155)	(0.109)	(0.0716)
PMT Mean	13.161	13.148	13.149	0.429	0.448	0.434	3.243	3.042	0.568	1.458	0.968	2.456
Observations	1121	1121	1176	1053	1051	1116	703	790	955	955	955	414
Panel B: Experimental Scale-up (PKH)												
Community Targeting	0.042	0.173*	0.100	-0.062*	-0.044	-0.087*	0.243***	0.235***	-0.025	-0.209	-0.045	0.313**
	(0.094)	(0.069)	(0.064)	(0.031)	(0.033)	(0.034)	(0.056)	(0.057)	(0.028)	(0.112)	(0.036)	(0.119)
PMT Mean	13.258	13.110	13.116	0.339	0.323	0.366	2.822	2.664	0.552	1.435	0.279	2.217
Observations	162	162	152	162	162	152	1074	1102	1835	1835	1835	279
Panel C: National Scale-up, BLT-DD only												
Community Targeting			0.126***									
			(0.004)									
PMT Mean			13.542									
Observations			17504									
Panel D: National Scale-up, Combined												
Community Targeting			0.054***									
			(0.002)									
PMT Mean			13.490									
Observations			41340									
Comparison Group												
Experimental Villages	×			×			×	×	×	×	×	×
Baseline Survey PMT		×	×		×	×						
Get Benefit Shares	Unequal	Unequal	Equal	Unequal	Unequal	Equal	Unequal	Unequal	Unequal	Unequal	Unequal	Unequal

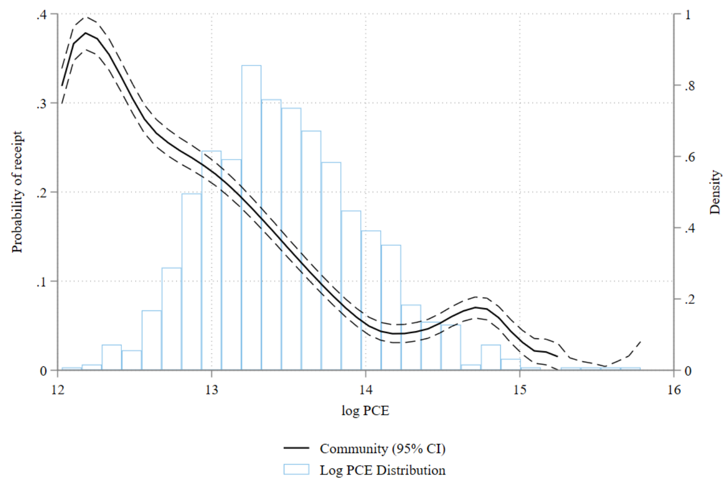
Notes: This table compares the per-capita consumption of program recipients by treatment arm. Standard errors are clustered at the village and are in parentheses. Pilot and PKH regressions include stratum fixed effects, while BLT-DD regressions include village fixed effects. SE clustered at the village level. See Table 1 for additional notes on the data and the sample restrictions. Panel D keeps households that are beneficiaries of other programs (e.g. PKH, KKS) in the BLT-DD analysis. $* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$.

Figure 1: Probability of Program Receipt by Log Per-capita Expenditure (PCE)

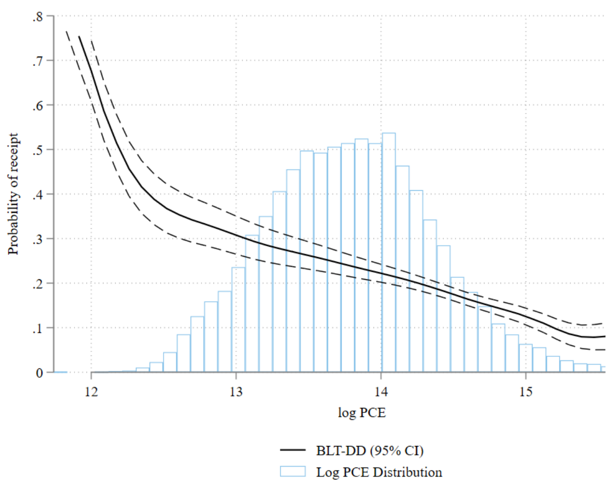
Panel A: Pilot (BLT)



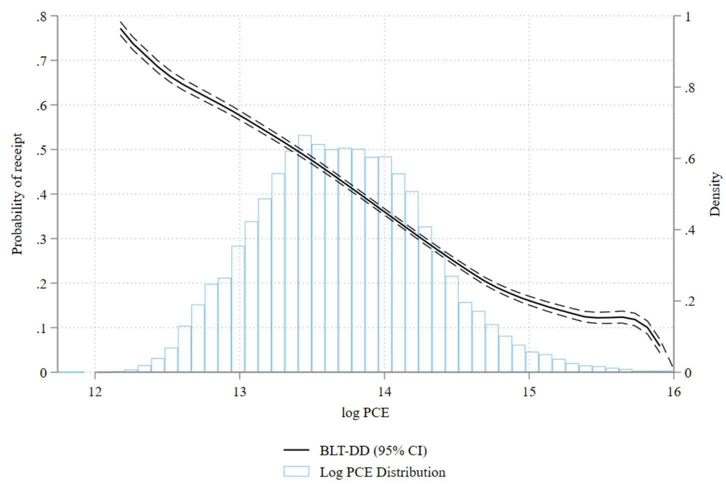
Panel B: Experimental Scale-up (PKH)



Panel C: National Scale-up, BLT-DD Only



Panel D: National Scale-up, Combined



Notes: Figure 1 shows the non-parametrical probability distribution (black) of receipt by log PCE and its 95% confidence interval (dashed lines) for the three programs (Pilot, PKH Expansion, and BLT-DD). For context, we also plot the log PCE histogram distribution (blue) in the background. Panel D keeps households that are beneficiaries of other programs (e.g. PKH, KKS) in the BLT-DD analysis. See Table 1 for additional notes on the data and the sample restrictions.

Online Appendix: Not for Publication

Appendix Table 1: Determinants of Benefit Receipt

	Pilot	*Experimental Scale-up*	*National Scale-up*	
	(BLT)	*(PKH)*	*(BLT-DD Only)*	*(Combined)*
	(1)	(2)	(3)	(4)
log PCE, 2020 IRP	-0.208***	-0.132***	-0.082***	-0.202***
	(0.017)	(0.022)	(0.004)	(0.004)
Mean if no receipt	13.638	13.533	13.842	13.842
Observations	1882	966	36355	48188
Below poverty line	0.245***	0.124***	0.067***	0.159***
	(0.034)	(0.041)	(0.009)	(0.008)
Mean if no receipt	0.101	0.119	0.071	0.071
Observations	1881	966	36355	48188

Notes: This Table replicates Table 2 using OLS instead of Conditional Logit. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 2: Determinants of Benefit Receipt, By Sub-Treatments and Removing Sample Restrictions

	Pilot (BLT)			*Experimental Scale-up (PKH)*				
				One In One Out			*Addition*	
	All	Elite	Full-Community	All	Elite	Full-Community	All	No Pre-Selected
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: log PCE								
log PCE, 2020 IRP	-1.287***	-1.442***	-1.153***	-1.662***	-1.316***	-2.122***	-1.453***	-1.321***
	(0.116)	(0.180)	(0.149)	(0.282)	(0.366)	(0.432)	(0.216)	(0.224)
Mean if no receipt	13.640	13.675	13.604	13.527	13.512	13.541	13.601	13.608
Observations	1667	819	848	655	317	338	552	482
Panel B: Below poverty line								
Below poverty line	1.145***	1.321***	0.979***	1.092***	0.833*	1.414***	0.839**	0.797**
	(0.163)	(0.262)	(0.196)	(0.313)	(0.483)	(0.399)	(0.327)	(0.367)
Mean if no receipt	0.106	0.088	0.124	0.100	0.123	0.078	0.119	0.118
Observations	1667	819	848	655	317	338	552	482

Notes: See Table 2 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 3: Determinants of Benefit Receipt, By Java and non-Java

	Java				*Non-Java*			
	Pilot	*Experimental Scale-up*	*National Scale-up*		*Pilot*	*Experimental Scale-up*	*National Scale-up*	
	(BLT)	*(PKH)*	*(BLT-DD Only)*	*(Combined)*	*(BLT)*	*(PKH)*	*(BLT-DD Only)*	*(Combined)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel B: Below poverty line								
log PCE, 2020 IRP	-1.225***	-0.877**	-0.502***	-1.123***	-1.355***	-2.175***	-0.774***	-1.207***
	(0.152)	(0.372)	(0.065)	(0.044)	(0.177)	(0.375)	(0.044)	(0.037)
Mean if no receipt	13.670	13.430	13.781	13.760	13.614	13.590	13.865	13.859
Observations	786	247	6017	14383	881	408	16052	27549
Panel B: Below poverty line								
Below poverty line	1.041***	0.118	0.284**	0.941***	1.272***	1.858***	0.566***	0.770***
	(0.227)	(0.564)	(0.129)	(0.070)	(0.232)	(0.383)	(0.075)	(0.056)
Mean if no receipt	0.128	0.150	0.073	0.075	0.087	0.067	0.063	0.065
Observations	786	247	6017	14383	881	408	16052	27549

Notes: The table divides the Table 2 analysis of the three programs into Java and non-Java. Standard errors are clustered at the village level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 4: Relative Importance of Baseline Household Characteristics in Benefit Receipt (OLS)

	Pilot	*Experimental Scale-up*	*National Scale-up*	
	(BLT)	*(PKH)*	*(BLT-DD Only)*	*(Combined)*
	(1)	(2)	(3)	(4)
log PCE, 2020 IRP	-0.122***	-0.056**	-0.062***	-0.145***
	(0.021)	(0.025)	(0.006)	(0.006)
Share PCE to food	0.297***	0.079	0.084***	0.315***
	(0.080)	(0.104)	(0.023)	(0.023)
Household size	-0.025***	0.017*	-0.004**	0.013***
	(0.006)	(0.009)	(0.002)	(0.002)
Share children	-0.011	0.305***	-0.019	-0.083***
	(0.053)	(0.081)	(0.012)	(0.012)
HH head with primary education or less	0.003	0.039	0.029***	0.090***
	(0.025)	(0.026)	(0.004)	(0.005)
HH head from other province	-0.009	-0.008	-0.004	-0.013
	(0.038)	(0.027)	(0.008)	(0.008)
HH head from other district	0.002	-0.034	-0.007	-0.017*
	(0.025)	(0.035)	(0.008)	(0.008)
HH head immigrated <5 years ago	0.063	-0.073*	0.016	-0.039***
	(0.039)	(0.040)	(0.013)	(0.013)
Widowed	0.123***	0.043	0.039***	0.037***
	(0.027)	(0.030)	(0.006)	(0.006)
Disability	-0.033	0.084*	0.014	0.016*
	(0.035)	(0.048)	(0.009)	(0.008)
Sick	0.006	0.047	0.017***	0.018***
	(0.029)	(0.032)	(0.005)	(0.005)
Has savings account	-0.078***	-0.012	-0.058***	0.053***
	(0.024)	(0.027)	(0.006)	(0.006)
Has motor vehicle	-0.220***	-0.103***	-0.007	-0.053***
	(0.023)	(0.028)	(0.007)	(0.007)
Remittances main source of income	-0.057	-0.053	0.022**	0.042***
	(0.042)	(0.041)	(0.010)	(0.010)
Temporarily unemployed	0.063*	0.018	0.019	0.015
	(0.036)	(0.040)	(0.013)	(0.014)
Avg. receipt prob.	0.312	0.148	0.184	0.184
Observations	1881	966	36355	48188

Notes: This Table replicates Table 3 using OLS instead of Conditional Logit. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

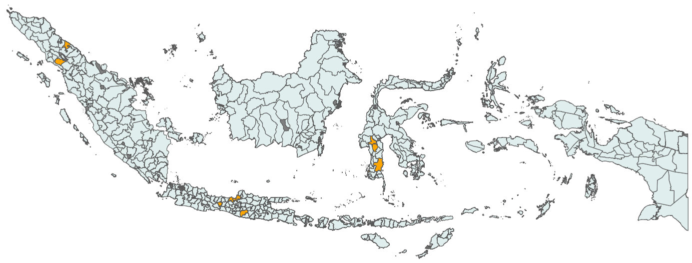
Appendix Table 5: P(Get Program) by Community Rank Percentile × Treatment Arm

	PCE Percentile			Community Rank Percentile		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pilot (BLT)						
Rank × Community Targeting	0.060	0.082	0.123***	-0.430***	-0.535***	-0.404***
	(0.048)	(0.049)	(0.037)	(0.057)	(0.058)	(0.044)
PMT Mean	0.295	0.295	0.306	0.299	0.298	0.309
Observations	3803	3803	3848	3527	3527	3608
Panel B: Experimental Scale-up (PKH)						
Rank × Community Targeting	-0.042	0.013	0.092*	-0.192***	-0.179***	-0.124**
	(0.036)	(0.037)	(0.039)	(0.052)	(0.052)	(0.045)
PMT Mean	0.054	0.054	0.076	0.054	0.054	0.076
Observations	2996	2996	2000	2995	2995	2000
Panel C: National Scale-up, BLT-DD Only						
Rank × Community Targeting			0.249***			
			(0.009)			
PMT Mean			0.184			
Observations			72790			
Panel D: National Scale-up, Combined						
Rank × Community Targeting			0.188***			
			(0.006)			
PMT Mean			0.380			
Observations			96382			
Comparison Group						
Experimental Villages	×			×		
Baseline Survey PMT		×	×		×	×
Get Benefit Shares	Unequal	Unequal	Equal	Unequal	Unequal	Equal

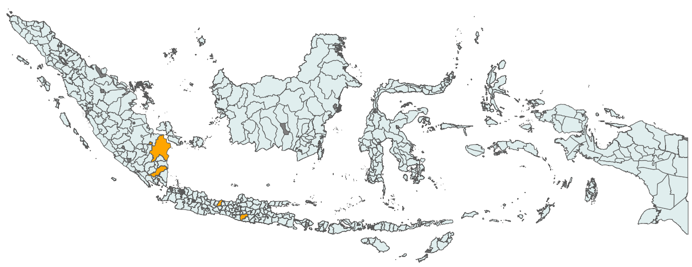
Notes: This table presents the results of an OLS regression of program receipt on (within-village) per-capita consumption percentile (columns 1-3) and community income-ranking percentile (columns 4-6), interacted with targeting-method. Pilot and PKH regressions include stratum fixed effects, while BLT-DD regressions include village fixed effects. See Table 1 for additional notes on the data and the sample restrictions. Standard errors are clustered at the village level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Figure 1: Program Districts

Panel A: Pilot



Panel B: Experimental Scale-up (PKH)

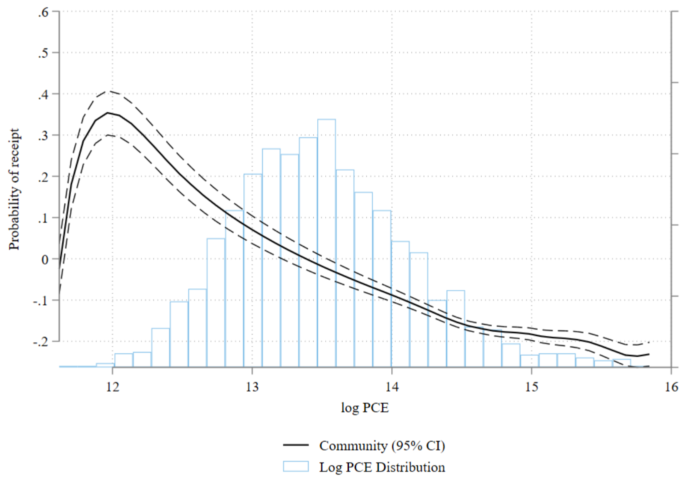


Panel C: National Scale-up, BLT-DD Only

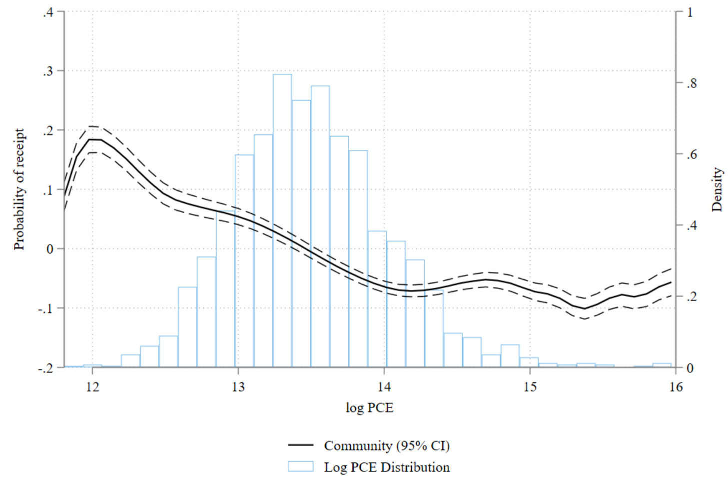
Notes: Figure 1 illustrates the geographic distribution of the Pilot Experiment, PKH Expansion Experiment, and the scale-up of the BLT-Dana Desa Program across Indonesia.

Appendix Figure 2: Probability of Program Receipt by Log Per-capita Expenditure (PCE), Residualized by Village FE

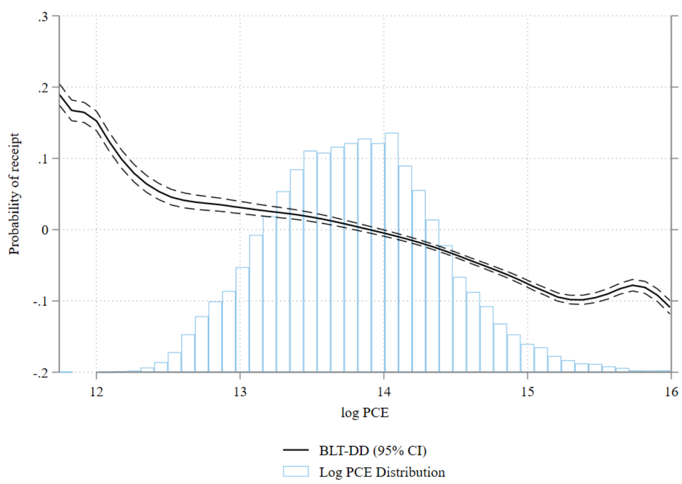
Panel A: Pilot (BLT)



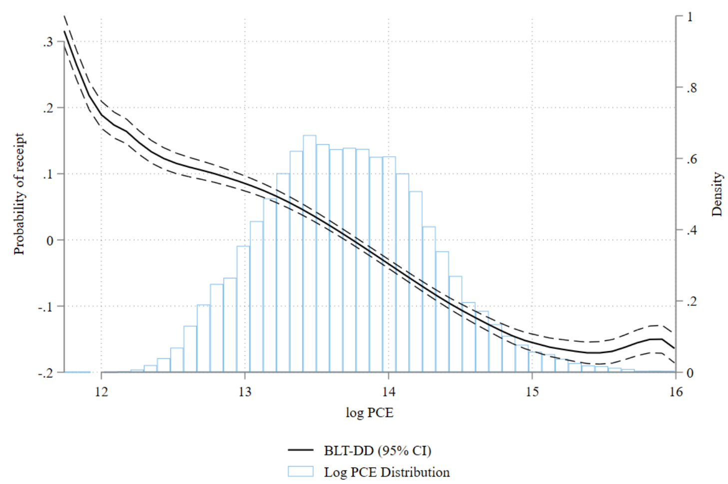
Panel B: Experimental Scale-up (PKH)



Panel C: National Scale-up, BLT-DD Only

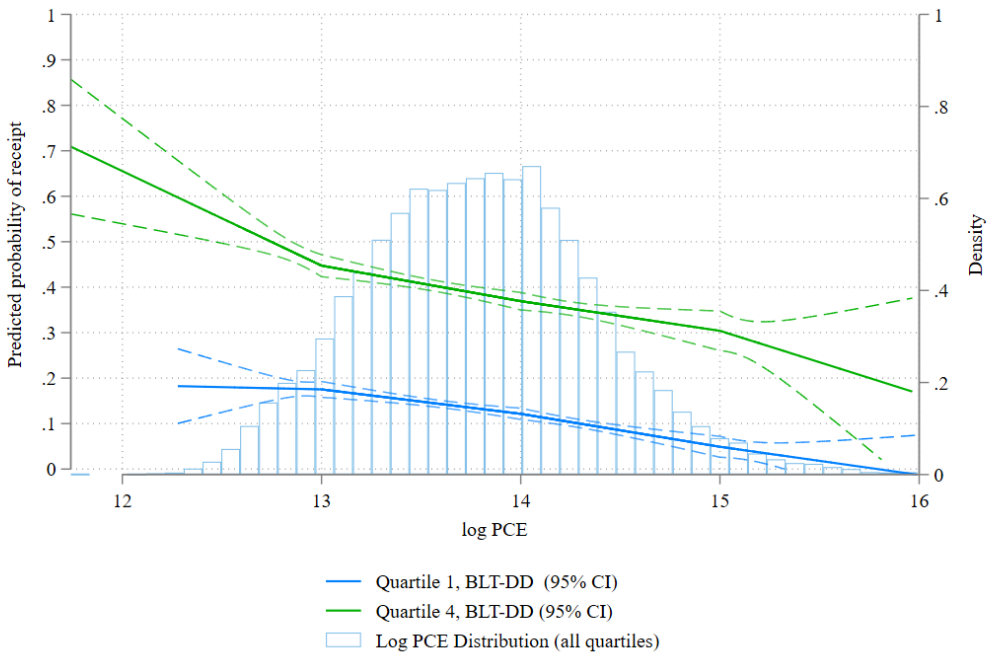


Panel D: National Scale-up, Combined



Notes: Figure 1 shows the non-parametrical probability distribution (black) of receipt by log PCE and its 95% confidence interval (dashed lines) for the three programs (Pilot, PKH Expansion, and BLT-DD). For context, we also plot the log PCE histogram distribution (blue) in the background. See Table 1 for additional notes on the data and the sample restrictions.

Appendix Figure 3: Predicted Probability of Program Receipt by Log Per-capita Expenditure (PCE), Comparing the First and Fourth Quartiles of BLT-DD Receipt



Notes: This figure presents a comparative analysis of the predicted probability of receiving the BTL-DD cash transfer between desas with the highest (Quartile 4) and lowest (Quartile 1) population densities. To get the predicted probabilities we employed an OLS regression where the dependent variable is a binary indicator of BTL-DD receipt, and the independent variable is the log PCE, represented through a linear spline with knots at equally spaced intervals (13-14, 14-15, 15-16, 16+).