Bulletin of Indonesian Economic Studies

ISSN: 0007-4918 (Print) 1472-7234 (Online) Journal homepage: https://www.tandfonline.com/loi/cbie20

Designing Anti-Poverty Programs in Emerging Economies in the 21st Century: Lessons from Indonesia for the World

Benjamin A. Olken

To cite this article: Benjamin A. Olken (2019) Designing Anti-Poverty Programs in Emerging Economies in the 21st Century: Lessons from Indonesia for the World, Bulletin of Indonesian Economic Studies, 55:3, 319-339, DOI: 10.1080/00074918.2019.1690411

To link to this article: https://doi.org/10.1080/00074918.2019.1690411

Published online: 18 Nov 2019.

Routledge
Taylor & Francis Group
Bulletin of Indonesian Economic Studies, Vol. 55, No. 3, 2019: 319–339

Indonesia in Comparative Perspective

DESIGNING ANTI-POVERTY PROGRAMS IN EMERGING ECONOMIES IN THE 21ST CENTURY: LESSONS FROM INDONESIA FOR THE WORLD

Benjamin A. Olken*
Massachusetts Institute of Technology

Governments of developing countries around the world have dramatically expanded social protection programs for the poor in recent decades. In doing so, they face a host of challenges in the targeting, design and implementation of these programs. In this paper, I describe the results from more than a decade of collaboration with the Indonesian government to understand how best to tackle these challenges, drawing primarily on evidence from randomised controlled trials. I highlight results that show the advantages of both community-based targeting and self-targeting, the importance of tangible information about beneficiaries' rights in minimising leakage, and the remarkable impacts of conditional cash transfers in the medium term. I also describe several recent studies that use randomisation at scale to generate policy-relevant evidence.

INTRODUCTION

Governments of developing countries around the world have dramatically expanded social protection programs for the poor in recent decades: between 2000 and 2017, the number of developing countries with at least some type of social safety net program expanded from 72 to 149 countries (World Bank 2017). As a result, most countries now have some type of social protection program. Conditional cash transfers (CCTs) are a particularly popular type of program, which typically give cash to poor households that meet some conditions, usually related to child health and education. More than 60 countries have some type of CCT program, up from 27 in 2008, and more than 90 countries have some type of unconditional cash transfer (UCT) program (Honorati, Gentilini and Yemtsov 2015).

Of course, helping poor citizens is not a new idea in developing countries. But there has been a global shift over the past two or three decades, away from subsidising basic commodities, such as food or energy, and towards more targeted programs, where aid is given only to households deemed eligible. One reason for this shift is that subsidies, while simple and transparent, may not actually have a

* Contact email: bolken@mit.edu.

ISSN 0007-4918 print/ISSN 1472-7234 online/18/000319-339
http://dx.doi.org/10.1080/00074918.2019.1690411

large impact on poverty. For example, if a country subsidises fuel, but consumption of fuel is proportional to income, then most of the benefits of the subsidy will end up going to the middle or upper classes, rather than to the poor. While it is possible to find amenities that are consumed mainly by the poor, generally the middle and upper classes consume more of most commodities. Thus, subsidising basic commodities is generally less efficient than targeted transfer programs for reducing poverty. One reason this shift is happening now, rather than, say, 30 or 40 years ago, is that while implementing targeted transfer programs is logistically more complicated than subsidising commodities, recent developments in information technology mean that these implementation challenges can more easily be addressed. For example, targeted programs require keeping track of a country's population, figuring out who is poor and making sure that people do not double-collect—i.e. receive the transfer more times than they are entitled to.

The evolving situation in Indonesia mirrors these global trends. Over the past two decades, Indonesia has implemented a number of social protection programs targeted directly towards poor households, including rice subsidy programs, Rastra/the BPNT, and their predecessors, Raskin and the OPK; direct cash transfer programs, the BLT and BLSM; and CCT programs, the PKH and BSM, among others.[1] The introduction of these programs has coincided with a scaling back of blanket subsidies, particularly of energy, in Indonesia. The trend away from across-the-board energy subsidies towards targeted transfer programs improves the ability of the government to ensure assistance reaches those who need it the most. This shift also reduces the distortions associated with social protection programs, since cash grants are less distortionary than subsidies, which change the marginal price of a good.

Implementing targeted transfers, however, creates new policy challenges. If we want to give assistance only to poor households, how do we identify which households are poor and should be eligible? What type of assistance should we provide: cash, in-kind transfers or something else? Concerning implementation and governance, how do we ensure that eligible beneficiaries actually receive the transfers they are entitled to? And finally, when looking at the bigger picture, does any of this matter? Are these programs effective in reducing poverty and improving well-being?

In this paper, I will focus on identifying those challenges, sharing findings from some of the work I have done in Indonesia, with a large number of collaborators, in order to shed light on different aspects of these questions. In particular, I will draw out some of the lessons that we have learned from decades of work with the Indonesian government to bring evidence to bear on these issues, for the future of anti-poverty programs both in Indonesia and in other emerging economies facing similar challenges.

The answers to these questions will be very different in a developing country context, such as in the context of Indonesia, than in a developed country context.

1. Rastra (Rice for the Prosperous Population program); BPNT (e-voucher, non-cash component of Rastra); Raskin (Rice for the Poor program); OPK (Special Market Operation program); BLT (Direct Cash Assistance program); BLSM (Temporary Direct Community Assistance program); PKH (Hopeful Families Program); BSM (Poor Students Assistance program). See glossary for the Indonesian names of these programs.

For example, for the question of targeting—in developed countries, the government typically can observe almost everyone's income, so it can target transfers to people with low income. Granted, targeting by income may not be a perfect solution (because it essentially taxes earnings, thus potentially discouraging work; see Moffitt 2002; Krueger and Meyer 2002). Nevertheless, targeting by income is typically not a viable solution in a developing country because governments often do not observe income, particularly for the poor, as most of the poor are out of the tax net (Jensen 2019). Instead, other approaches need to be designed so that they are customised to the realities of the developing country context. But these alternative approaches involve trade-offs, and it is an empirical question as to which methods work best. The answer may also depend on how to define what 'best' means in a particular context.

In terms of what kind of assistance to provide—cash, in-kind transfers or something else—the answer to this question will also differ between contexts. For example, there is less banking infrastructure in many developing countries, which may make the distribution of cash more difficult. Supply issues may also be important to consider. If a government provides subsidies in remote areas, but the subsidised good is not readily available in those areas, then subsidies may affect local prices (Cunha, De Giorgi and Jayachandran 2019).

Challenges of program implementation, governance and leakages may also be particularly severe in many developing countries (Olken 2006). Transparency and access to information may be challenges as well. For example, developing countries may experience relatively low levels of literacy, particularly among poor beneficiaries, who may not understand their rights under a given program.

Finally, we also need to understand the impact of these types of social protection programs on poverty and well-being. While a lot of research has been conducted to understand the impact of these programs in the short run, understanding their impact over the longer term is more difficult and an important area for new research.

For the past decade in Indonesia, my colleagues and I have tried to address these questions in a developing country context. Throughout this process, we have collaborated with many government colleagues and agencies in Indonesia, including the National Development Planning Agency (Bappenas), Statistics Indonesia (BPS), the Ministry of Social Affairs (Kemensos) and the National Team for the Acceleration of Poverty Reduction (TNP2K) under the vice president's office. Through this unique partnership, we have worked together to conduct a number of randomised policy evaluations—the gold standard of evidence. By randomising which locations get which treatment, we can truly understand the impacts of particular policies. This article describes our findings, the implications for current and future anti-poverty programs in Indonesia and the implications for other countries facing similar challenges.

TARGETING: HOW DO YOU IDENTIFY WHO IS POOR?

Targeting Methods

There are three main approaches to thinking about poverty targeting in a developing country context. The first approach is a proxy means test (PMT). In a traditional means test, if a person's income is below a given level, they are eligible for the program. When a government cannot observe income, it can instead use a PMT, which predicts a person's income based on variables that can be observed, creating proxies for income. The government can then determine eligibility based on this predicted income. In other words, a PMT is based on proxies of income rather than actual income.

To conduct a PMT, government enumerators usually go from door to door and conduct a census of all potentially poor households' observable assets, such as house-building materials, vehicles, televisions, etc. Government analysts then run a regression in a different dataset, collected for research purposes, that estimates the relationship between the assets they can observe and the income or consumption level of a household, which is observed in research data but not observed for the entire population. Once they have that regression, they can return to their census assets data and predict a household's income. As such, a PMT typically involves a large data collection effort. In Indonesia, for example, a PMT census has occurred about once every three years since 2005 through the Social Protection Registration (PPLS) survey. Depending on the year, the government enumerators will visit between 15 and 20 million households to collect data on different asset characteristics.

Moreover, the proxies created for income may be imperfect. For example, people have different tastes and consumption patterns. Given the same level of income, some people may prefer to spend money on a house while other people may prefer to save money in a bank account. A PMT will estimate that people who prefer to spend money on a house are richer than those who prefer to save money in a bank, because bank savings are not observed in a PMT. A PMT may also not reflect income shocks. For example, if someone is laid off from their job, they may still own the same house and assets, but their actual income may be quite low.

None of this is to say that a PMT is not valid. Rather, these observations highlight that the proxies are imperfect. At a technical level, one can measure the predictive power of a PMT through the R^2, which captures how well the variables in the PMT predict actual per capita consumption. In examples we have produced, these R^2 values are usually between 0.4 and 0.6, confirming that the PMT asset variables have substantial predictive power, but that there is still substantial residual variation in per capita consumption that is not captured.

A second approach is community-based targeting, which is based on the idea that communities may have better information than the government about which households are the poorest. However, if the government asks a community for this information, and links money to citizens identified as poor, elite capture may be a concern. For example, a commonly cited concern is that local elites may try to manipulate the process by putting their nephews or cousins on the eligibility list instead of the people that are the most deserving. How the community defines poverty may also be different from how the central government defines it; I will return to this issue in more detail later.

A third approach is 'self-targeting', where poor households are asked to apply for a program. The PMT that I described above entailed automatic enrolment—government enumerators went from door to door to screen households and placed households below a certain poverty threshold on the eligibility list automatically, without the households needing to take any action. In a self-targeting approach, by contrast, the government *first* asks anyone who is interested in the program to apply. If people apply, they are then screened with a PMT in order to determine whether they are eligible.

There are potential advantages and disadvantages to a self-targeting approach. In automatic enrolment, the government tries to identify all potentially poor households, but may miss a few, particularly if poor households live on the margins of villages. With a self-targeting approach, the poor can make themselves known, potentially reducing so-called 'exclusion error'. This can be particularly useful if the on-demand application process is open continually; households who were better off, but then experienced a negative economic shock, could potentially apply under an on-demand system. Under self-targeting, the upper or middle classes can also screen themselves out: if someone is in a relatively well-off household, they may not bother to apply for the program because they do not think they will pass the PMT screening test. This type of self-selection can help the government reduce so-called 'inclusion error' in the PMT formula. As discussed, the household that chooses to keep its money in a bank rather than spend it on building a house would perhaps pass a PMT test, but the household may not bother to apply for the program if it does not understand the screening formula.

However, a disadvantage of self-targeting is that some poor households may not apply. They may feel intimidated by the application process, or they may face time constraints. For example, to apply for the program, a poor household may need to take time off work to visit an application office, which may mean they do not earn any money that day and may have to skip meals in order to apply.

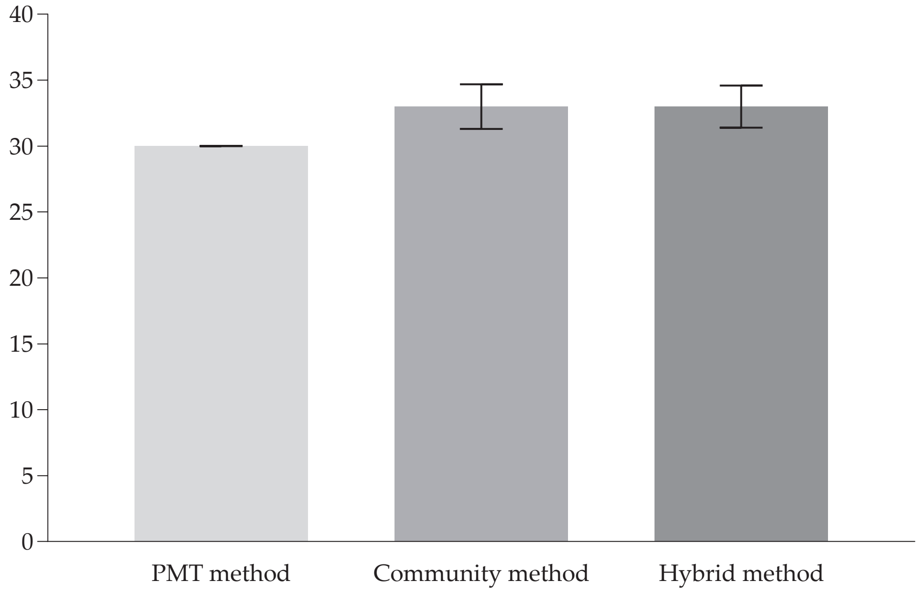
Comparing Targeting Methods Using Randomised Trials

It is not obvious *ex ante* which targeting method is most effective, as they all involve trade-offs. Hence, we conducted two different randomised evaluations to understand, in an empirical manner, which of these targeting methods was the most effective. In the first study, done with Alatas et al. (2012), we compared a PMT approach with a community-based targeting approach and a hybrid method for a one-time cash transfer of about $3. In the second evaluation, done with Alatas et al. (2016), we conducted a larger experiment in the context of the expansion of the PKH. We compared an automatic enrolment PMT with a self-targeting-based PMT, which required households to apply before receiving the same means test.

Comparing PMT with Community Targeting

In the first evaluation, we randomly assigned 640 hamlets, or sub-villages, to one of three targeting methods: a PMT, community targeting or a hybrid method (note that this experiment included both rural *desa* and urban *kelurahan*; I will henceforth refer to both of them as villages for simplicity). In the PMT, we used 49 indicators created by BPS, using its normal PMT method. In the community method, we facilitated community meetings in which communities ranked households from richest to poorest. Communities first discussed what their view of poverty was, and then community members took turns ranking everyone in the community from richest to poorest, placing name cards on a string. This produced a complete rank ordering of the relative perceived poverty levels of everyone in the neighbourhood. The households ranked the poorest received the transfer. The third method was a hybrid, where we first conducted the community-based approach and then applied the PMT in order to screen out incompatibility. In conjunction, we also conducted

FIGURE 1 *Targeting Error under Each Targeting Method (%)*



Note: Results show targeting error, defined as either inclusion or exclusion error, and error bars represent 95% confidence intervals from Alatas et al. (2012).

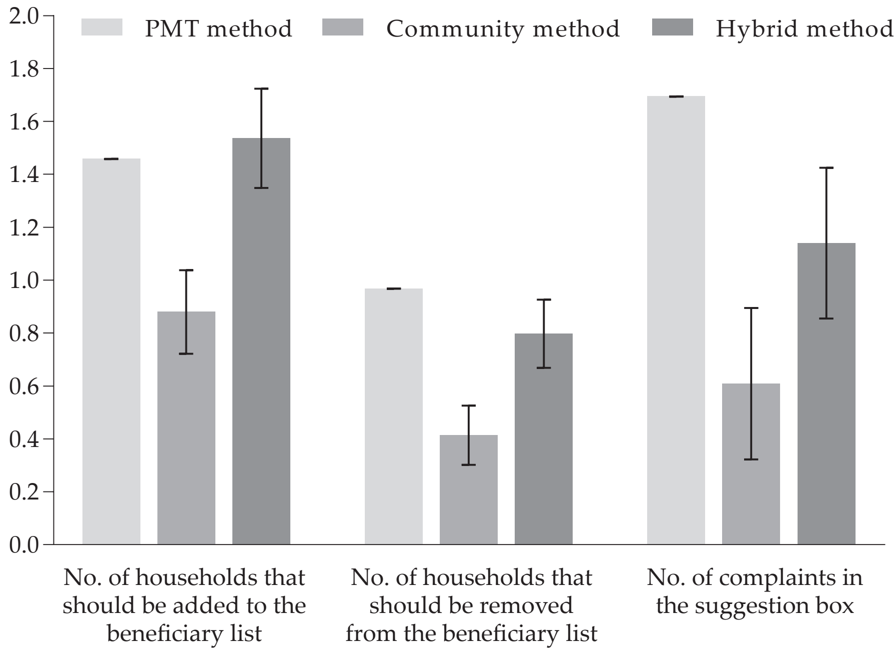
an independent survey to determine a household's true poverty level in order to measure which of these targeting methods was most effective at identifying poor households.

To assess outcomes, we looked at both targeting error and community satisfaction. Targeting error was defined as either giving the transfer to a non-poor household or failing to give the transfer to a poor household. We measured whether a household was actually poor by conducting a consumption survey *before* the targeting process and by matching the results of that consumption survey to the targeting results. We measured community satisfaction in a variety of ways in an end-line survey and also established a comment box where people could leave feedback about the targeting method used.

The results in figure 1 indicate that the PMT method had less targeting error than the other two methods by about 3 percentage points (10%). If we look at the data more closely, the targeting methods mostly appear to agree on which households are the rich and which are the very poor. Where the PMT does a slightly better job of sorting households is with people very close to the threshold. Thus, the overall impact on poverty from using one method versus another actually looks very similar. In cases where the methods disagree, the differences in income are very small. In other words, even though the PMT technically does a somewhat better job of reducing targeting error, our calculations suggest that the PMT would not reduce poverty substantially more than the other targeting methods if it were scaled up.

We also asked households which targeting method they preferred and how well they thought each method worked. Did they think the method was reasonable?

FIGURE 2 *Impact of Three Different Targeting Methods on Community Satisfaction*



Note: Results show measures of community satisfaction, and error bars represent 95% confidence intervals from Alatas et al. (2012).

Would they prefer to use this method again if there were another targeting project in their village? We also showed participants the final list of eligible beneficiaries and asked if there was anyone they thought should be added to or removed from the list.

The results in figure 2 indicate that community targeting led to much higher community satisfaction and better-selected households that self-identified as poor. Under the community targeting method, the community wanted to make relatively fewer changes to the beneficiary list and also submitted fewer complaints, about a third lower than with the PMT.

The reason for these differences was not that the community performed worse or did not have local information. Rather, it appears that the community had more information than the central government and had chosen to weight this information differently. In other words, the community's definition of poverty differed slightly from the government's definition, which was based on per capita consumption. For example, widows self-identified as poorer than other people with the same per capita consumption level, perhaps reflecting a lower earning ability in their households. Communities also agreed and included widow-headed households on their targeting lists at a higher rate.

Not surprisingly, if the community has a different notion of what poverty is, and it targets based on its own definition, then it will also think it does a better job of targeting. Ultimately, the trade-offs between these two approaches come down to whether the government wants to target strictly based on per capita consumption, which is the typical measure used by the government to measure poverty, or whether it wants to recognise that communities have slightly different perceptions of poverty, and include those as well.

A second potential reason for the higher level of satisfaction with the community-based approach may be that it is a much more transparent method than a PMT. Because a PMT is based on complicated formulas, it is inherently less transparent than a community-based approach. For example, a household may have a television but also have a dirt floor and a thatched roof, while another household may have no television and a cement floor and a tin roof. Under a PMT test, a household likely has no idea why those variables are relevant and how they combine to determine eligibility, and hence why one household receives the program while the other does not. We also found no evidence of elite capture in the community-based approach.

Taken together, these results imply that there are important trade-offs to consider when choosing a targeting approach, and there can be benefits to adding a community-based approach to existing targeting methods.

Comparing Self-targeting with PMT

In the second targeting evaluation, we compared an automatic targeting approach with an on-demand application in the context of the expansion of the PKH. The PKH, Indonesia's CCT program, targets the very poorest people in Indonesia. Targeting for this program has high stakes, as households receive about 11% of their consumption, or about $900, over six years.

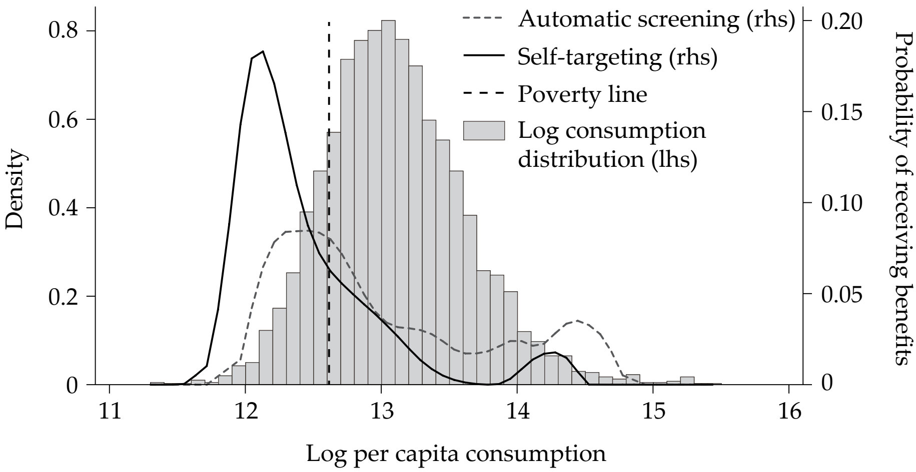
For the study, a facilitator visited each village and explained that the program was beginning and was going to screen for poverty, without giving the exact screening formula. Under the automatic PMT method, the government conducted a PMT from door to door and automatically enrolled households below the poverty threshold. Under the on-demand application method, households were required to apply for the program in advance. If people were interested in the program, they had to apply at a specific time and nearby location, either at the subdistrict (*kecamatan*) or the village (*desa/kelurahan*) office or in their hamlet (*dusun*). If they applied, their poverty status was verified by a PMT, and for those households close to the margin of eligibility, by a follow-up home visit as well.

The results in figure 3 indicate that requiring households to apply for benefits led more poor and fewer non-poor households to receive benefits compared with automatic screening. In other words, in on-demand-application villages a person had a higher probability of receiving benefits if they were very poor and a lower probability of receiving benefits if they were non-poor.

There are several different mechanisms driving these impacts. First, in the automatic enrolment villages, it is possible that, despite the best efforts of government enumerators, they may have missed some poor households during the PMT identification process. In comparison, in the on-demand approach, any poor household could make itself known and apply, leading more poor households to receive the program and to a reduction in exclusion error.

Second, wealthier, ineligible households appear to have chosen not to apply. This can have significant budget implications, because most of the population is ineligible (see the histograms in the background in figure 3). In general, since there are errors in the PMT, a small fraction (about 2%–3%) of households over the eligibility threshold end up on the eligibility list. While there is a very small

FIGURE 3 *Distribution of Beneficiaries under Self-targeting vs. Automatic Screening*



Notes: Author's calculations showing probability of receiving benefits by per capita consumption level and treatment, using data from Alatas et al. (2016). The *light-grey histogram* shows the overall distribution of log per capita consumption.

probability that these households receive the benefit, since 90% of the population is ineligible, the 2%–3% of ineligible households that pass the PMT screen by mistake actually make up a relatively large proportion of total beneficiaries. In comparison, under self-targeting, many of those ineligible households choose not to apply. In particular, under self-targeting, 61% of eligible households apply, while only 10% of ineligible households apply, indicating that households take their own income status into account when deciding whether to apply.

These two effects combined meant that the beneficiaries who were selected were 20% poorer on average in the villages with on-demand application than in the villages with automatic enrolment. On-demand application, even though it has some potential risks, seems practical and effectively reduces both inclusion and exclusion errors.

One important issue is that different targeting methods may work better in different types of areas. To test this, both our targeting studies—comparing PMT with both the community-based approach and the on-demand application approach—took place in a mix of urban/peri-urban and rural environments. In most cases, the results are broadly similar in these environments. Although the differences between results from both PMT and the community approach and PMT and the on-demand approach are somewhat less pronounced in urban areas, the differences between urban and rural areas are not statistically significant in most cases.

The (Lack of) Distortionary Effects from PMT Targeting

A third question we examined is whether knowledge of the variables used to determine eligibility for benefits under the PMT affects households' investment decisions. In other words, does telling people about the assets that determine their eligibility for benefits distort their investment decisions?

To investigate this question, my colleagues and I have been working with BPS to conduct a randomised evaluation, using a survey in 2015 for the national database for social assistance (PBDT) (Banerjee, Hanna, Olken and Sumarto 2018). In this survey, the government randomly varied the screening questions asked in different provinces. To keep the number of questions in the survey constant, each randomised question had one of two options. In half of the provinces, households received (1) either a question on flat-screen television ownership or a question on the number of rooms in the house (randomised by province), and (2) either a question on the number of active mobile phone SIM card numbers in the house or on whether the house had a modern toilet installed (also randomised by province). We specifically chose these two key treatment questions—on flat-screen television and SIM card ownership—because we had access to independent data sources on actual asset ownership that did not rely on household self-reports.

We found that people appeared to be paying attention, but the distortionary effects seemed limited. In provinces where the PBDT questionnaire asked about television ownership, households were about 16% less likely to report owning a flat-screen television in the next round of Susenas, carried out six months later. However, by the next year, this effect had disappeared. More importantly, *actual* television sales—as measured by an independent survey of television retailers—did not appear to decline. We also saw no changes in any of the other assets we tracked. Thus, while there is some evidence that asking those asset questions may change how people strategically respond in a government survey, it does not seem to actually change consumption behaviour.

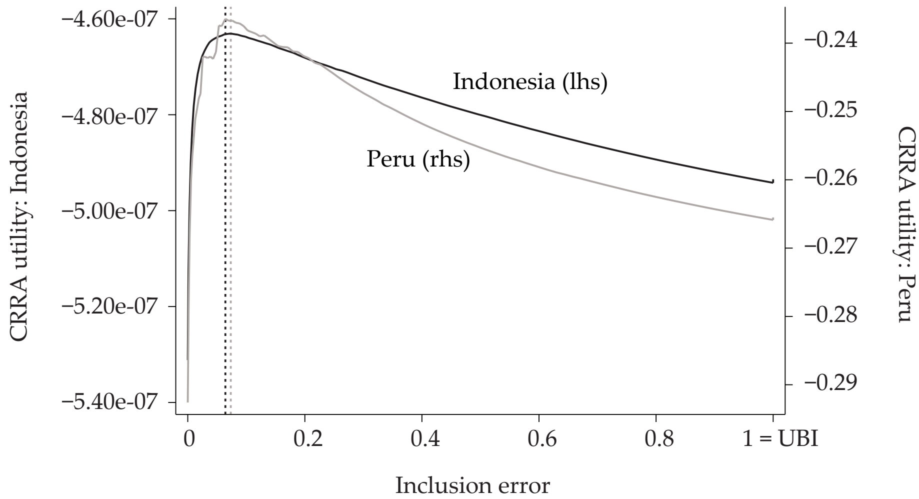
Is Targeting Worth It?

The final topic I discuss in this section is the decision about whether to target at all. There has been tremendous attention in the world recently on a universal basic income (UBI). The general idea of a UBI is that instead of trying to figure out who is eligible, the government simply gives the transfer to everybody. Such a scheme is still redistributive: if a country taxes in proportion to consumption, the rich are still paying substantially more than the poor. So, if everyone is receiving the same transfer, the net effect of the scheme is to redistribute towards the poor.

The main downside of a universal transfer is that the beneficiaries receive less per person. For a given budget, if the government targets only 10% of the population, it can give each beneficiary 10 times more than if it gives the transfer to 100% of the population. In other words, when countries have a fixed budget, the transfer per person mechanically decreases when more people are included. With a universal transfer, the government does save money because it does not have to run a poverty census, but it turns out that the cost of running a poverty census, even though it is large in dollar terms, tends to be very small compared with the budget of the transfer program itself.

Another benefit of a universal transfer is that the government saves the cost of targeting. While intuitively this saving may seem large—after all, government enumerators go from door to door to survey tens of millions of households—the saving pales in comparison to the magnitude of the benefits delivered through targeting. To be specific, the 2015 Indonesian poverty census cost the government about Rp 900 billion. The government typically conducts a new census once every three years; the annualised cost of targeting is therefore about Rp 300 billion. By

FIGURE 4 *Social Welfare vs. Inclusion Error Trade-off: Simulations for Indonesia and Peru*



Notes: Reproduced from Hanna and Olken (2018). For each cutoff value *c*, we calculate the per capita benefit amount for included households and then calculate the constant relative risk aversion (CRRA) utility with $\rho = 3$. If a household is not included in the program at a given value of *c*, it us assumed $b^i = 0$. *Dashed lines* indicate the point of maximum social welfare in each country. Inclusion error on the *x*-axis shows the inclusion error from each potential cutoff value *c*. UBI = universal basic income.

contrast, in 2019, total expenditures on targeted social protection expenditures (i.e. the value of all transfers allocated using the targeting census) amounted to Rp 98 trillion (TNP2K 2019). The annualised cost of running the targeting census, expressed as a fraction of the total amount of assistance being targeted, is therefore about three-tenths of 1%. Given this, the benefits of not having to run the targeting census tend to be swamped by other considerations.

To quantify the remaining trade-offs, Rema Hanna and I conducted a simulation where we considered different targeting approaches and their related targeting errors, holding the budget constant, and compared them with a universal program (Hanna and Olken 2018). How well a targeted program performs depends in part on the accuracy of the targeting. If the targeting is very accurate, then a targeted transfer will perform relatively well at improving social welfare. If the targeting is not very accurate, then a universal transfer might perform better.

Figure 4 shows the results for simulations we conducted using data from both Indonesia and Peru. The results indicate that a program that is relatively narrowly targeted (i.e. one that targets about 19% of the population, which is reasonably close to the current Indonesian national targeting policy for many programs) achieves a much higher level of social welfare than universal programs. This result holds even when accounting for the cost savings from not having to run a poverty census, which, as discussed above, turn out to be very small in comparison to the transfers given out. In other words, targeted transfers, while imperfect, deliver a far greater improvement in welfare than a UBI would, because they can transfer much larger amounts per beneficiary. The downside is that such a program does miss deserving people: in our calculations, exclusion error from such a program would be about 30% in Indonesia and about 20% in Peru.

In sum, findings from these studies on different targeting methods suggest that community feedback can improve targeting, in the sense that it can improve the perceived legitimacy of targeting lists, and citizens' satisfaction with them, by making them conform better to local perceptions of poverty. Adding on-demand application to a PMT can improve targeting, by reducing both inclusion and exclusion errors. We do not see any distortionary effects in household consumption or investment behaviour from the PMT surveys. And when countries have fixed budgets, targeted transfers—at least based on our simulations—look much more effective than universal transfers for helping the poor and improving social welfare.

WHAT TYPE OF ASSISTANCE SHOULD WE PROVIDE?

What type of assistance should we provide: cash, vouchers or in-kind transfers? With each type of transfer, there is a trade-off between placing restrictions on how transfers can be used versus giving people more choice. Transferring cash allows the most choice for individuals. In-kind transfers, where the government gives recipients the good itself, offer the most government control. Vouchers are an in-between option, where the government typically allows people to choose where to spend the transfer, but only for a restricted set of goods.

Each of these options has pros and cons. Cash is very flexible, but there are political concerns about how people will spend the money, and concerns that prices may rise if there is limited supply. In-kind transfers can improve the supply of the subsidised good, which can be helpful in a limited-supply environment, but they also raise concerns about quality and leakages. In addition, people may not need the particular good they are given and may value it less than cash, which they can use to buy what they need most. To the extent that the goal of the transfers (CCT programs, for example) is to stimulate health and educational investments, other questions arise. In particular, different approaches may be needed in remoter areas with less-developed supply. I will discuss some of Indonesia's pioneering work to adapt ideas about CCTs to these kinds of environments.

Some Global Evidence

There is some existing global evidence addressing these questions. For example, Cunha, De Giorgi and Jayachandran (2019) conducted a randomised trial in Mexico, where the government randomised whether beneficiaries in particular communities received in-kind transfers or cash. They found that in isolated communities, giving out cash caused prices for basic commodities to increase. The cash transfers created a demand shock and because of limited supply, caused prices to rise. In comparison, giving out in-kind assistance actually lowered prices in isolated communities, acting as a positive supply shock. In Indonesia, one can imagine that price effects would not be a large concern in areas with thick markets, such as Java. However, in remoter areas, price effects could be very important to consider. Overall, this evidence suggests that existing supply constraints should be considered when switching from in-kind transfers to cash transfers in isolated areas. We are exploring these questions in ongoing work in Indonesia.

A second study by Aker (2017) compared cash transfers with vouchers in a randomised evaluation in an internally displaced person (IDP) camp in the Democratic Republic of the Congo. The study found little difference in overall well-being under either method, in part because people purchased similar goods. In urban Ecuador, Hidrobo et al. (2014) conducted a randomised evaluation to compare cash with in-kind transfers and with vouchers and found that all three types of transfers improved the quantity and quality of food consumption. The in-kind food transfers increased calories the most, while the vouchers increased dietary diversity the most.

One other finding is worth noting. Across these studies, and in many other randomised evaluations on cash transfers, there is no evidence for the commonly cited political concern that people spend cash transfers on 'temptation goods', such as tobacco or alcohol.[2] However, there may be other important concerns about transferring cash. For example, the government may want beneficiaries to spend the cash on food or a specific good, or supply may be limited.

Conditional Community Transfers in Indonesia

An important idea of conditional cash transfers is that by incentivising health and education, they may stimulate demand for these services. But in areas that are supply constrained, stimulating demand for these services (which could lead to price increases if private providers are involved) may not necessarily increase actual service uptake. For remote locations, another approach may make sense.

Indonesia again was a leader worldwide in thinking about these issues. In 2005, Indonesia began planning the introduction of a CCT program in a working group led by Bambang Widyanto, then at Bappenas. Given the concerns raised above, Indonesia chose to pilot two different types of programs: a traditional, household-based CCT program, and a new program aimed at more-rural locations and that moved cash transfers to the community level rather than the individual level. The idea behind the community-level grants was to directly increase supply in more-rural locations—e.g. a village could use the funds to hire a midwife or to start a remote-school classroom (*kelas jauh*)—rather than to stimulate demand through household cash transfers. This program was launched in 2007 and called the National Community Empowerment Program for a Healthy and Smart Generation (PNPM Generasi Sehat dan Cerdas, also known as Generasi). The idea of incentives from CCTs was incorporated into Generasi. Each village received points for the various health and education services it obtained, such as for each child immunised, each mother who received maternal care, each child enrolled in school, and so on, following the same set of conditions used in the household CCT. Villages that earned more points in each *kecamatan* would get a larger block grant the next year.

Moreover, remarkably, the government chose to evaluate *both* programs using randomised controlled trials, at the *kecamatan* level. That is, the government first drew up a list of locations where the programs could take place. As funds were limited, not all *kecamatan* could receive the programs. The government therefore decided to randomly select which *kecamatan* would receive the program—176 *kecamatan* for Generasi and about 438 *kecamatan* for the PKH—and which *kecamatan*

2. See Aker (2017); Banerjee et al. (2015); Banerjee et al. (2017); Blattman, Jamison and Sheridan (2017); Evans and Popova (2016); and Haushofer and Shapiro (2016).

would serve as control groups. Each of these evaluations separately represented one of the largest randomised evaluations conducted anywhere in the world at the time. For Generasi, the government chose to evaluate both the program with incentives, as described above, and an identical program without incentives, in order to test if the incentives mattered. I worked most closely on the Generasi evaluation, along with Junko Onishi and Susan Wong from the World Bank, while Vivi Alatas from the World Bank led the team on the PKH evaluation.

The CCT program found positive results after two years—consistent with evidence from around the world (Alatas 2011). It is important to note that the hypothesised supply-side concerns were important. Even though the PKH was focused on 'supply-side-ready' locations, Triyana (2016) found that in the short run, Indonesia's CCT program (the PKH) increased the delivery fees charged by midwives. However, this did not prevent the program from increasing medical care overall. The number of midwives increased and so did the level of medical care given.

What about Generasi? We found that the program was effective in increasing the use of maternal and child health services, particularly weight checks for young children, and resulted in reductions in malnutrition (Olken, Onishi and Wong 2014). The program also increased school enrolments. The incentives in Generasi sped up how quickly the program results materialised. The program was most effective in the areas with the lowest levels of service delivery at baseline, such as rural locations in East Nusa Tenggara. Our calculations suggest that the cost-effectiveness of the program was similar to cost-effectiveness of the PKH, at least in the first two years.

These two evaluations suggest that 'one size may not fit all'. That is, different approaches may work best in different areas. Programs like Generasi may be particularly appropriate in areas with relatively low levels of service supply. In terms of policy impact, both programs, the PKH and Generasi, were subsequently scaled up. However, the scale-up of Generasi took this evidence in mind and focused primarily on the more rural and remote locations, where the program was found to be most effective.

How Do We Ensure that Assistance Reaches Eligible Families?

Like governments in many developing countries, the Indonesian government wants to ensure that all of the benefits from social protection programs actually reach the intended beneficiaries. However, challenges with leakages can reduce program impacts. How do we reduce leakages and ensure that all program benefits are reaching the targeted beneficiaries? In particular, can strengthening access to information improve service delivery?

To study this question, my colleagues and I conducted a randomised evaluation of Indonesia's Raskin/Rastra program and studied whether providing people with information on program benefits increased the amount of benefits that eligible households received (Banerjee, Hanna, Kyle, Olken and Sumarto 2018). Raskin, as it was called at the time, was a program where the poorest 30% of households were entitled to about 15 kg of rice per month at about one-fifth of the market price. However, there were substantial leakages in the implementation of this program, both with ineligible households receiving subsidised rice and with the actual delivery of the rice.

At the time we started, the government was considering introducing identification (ID) cards as tangible proof for beneficiaries of what their rights were under the program. Providing this information to citizens could both inform them of their eligibility and entitlements under the program and enable them to better bargain for their rights with local officials. However, it was unclear whether introducing the cards would actually work and be worth the potential economic and social costs. So, the government invited us to work with it to conduct a randomised policy evaluation to determine whether these ID cards would actually improve the functioning of the program.

In addition to testing the impact of the ID cards overall, the government wanted to know what variation of card would work best. We therefore evaluated several variations, such as cards containing information on just the quantity of subsidised rice that households were entitled to or cards containing both the quantity and price of the subsidised rice that households were entitled to; we also evaluated whether beneficiary lists should be posted publicly. We designed this evaluation explicitly in collaboration with the government to help inform its choice of program design.

Figure 5 indicates that the ID cards increased the subsidy that eligible households received by about 26%, relative to the subsidy received in comparison villages where ID cards were not mailed out. Interestingly, ineligible households received no less, indicating that there was an overall reduction in leakage from the program, rather than just a transfer from ineligible households to eligible ones.

The content of the cards also mattered. Conditional on sending out the cards, the impact of the subsidy was doubled in villages where the cards included both price and quantity information, compared with villages that received ID cards with only quantity information. Publicising beneficiary lists also increased the subsidy received by eligible households by about 25%, compared with villages that received ID cards without public information.

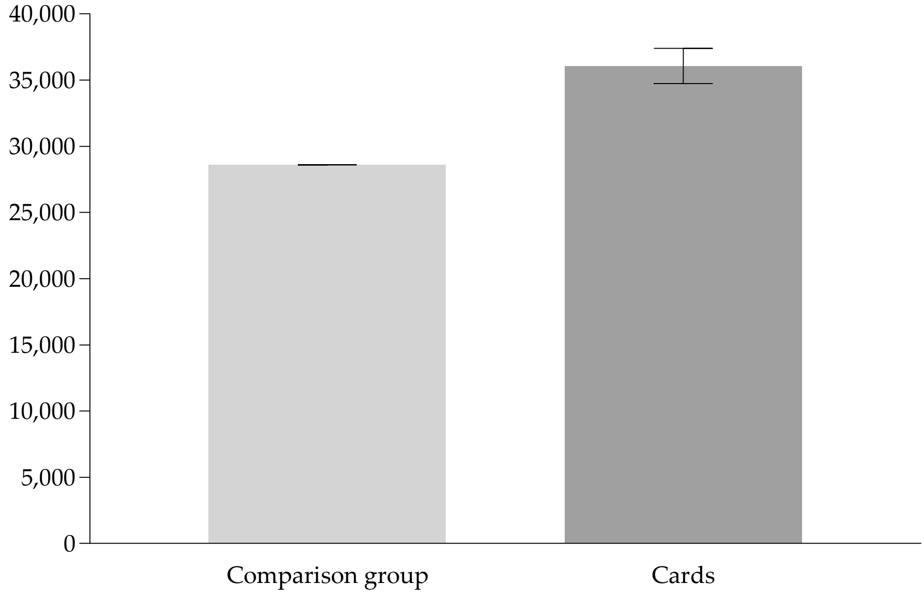
These results indicate that information matters, even for a program such as Raskin, which in various forms has been around since 1999. Giving people tangible information about their rights seems to be an important component in ensuring they receive what they are entitled to. These positive results were some of many factors that led the government to scale up this program. In 2013, the government provided social protection ID cards to 15.5 million poor households, reaching 65.67 million people. This is an example of how research can have concrete implications for policy, both in Indonesia and more generally.

WHAT IS THE IMPACT OF ASSISTANCE ON POVERTY AND WELL-BEING? EVIDENCE FROM CONDITIONAL CASH TRANSFERS

Global Evidence

What is the impact of transfer programs on poverty and well-being? I will focus here on CCTs, which, as described, are a common type of social protection program. CCTs are typically long-term programs where transfers are given regularly to poor households on condition that they make human capital investments in their children. For example, for expectant parents, the conditions may include prenatal care, delivery by a trained medical professional and so on. After the child is born, the conditions may include early childhood health investments and enrolment in

FIGURE 5 *Impact of Raskin Cards on Subsidy Received per Eligible Household (Rp/month)*



Notes: Results show the average amount of benefits received by eligible households in treatment and control villages, and error bars represent 95% confidence intervals from Banerjee, Hanna, Kyle, Olken and Sumarto (2018).

primary and secondary school. CCTs generally aim to improve the welfare of not only the recipients but also their children, by improving the children's human capital accumulation.

CCTs began in the 1990s in Mexico, Bangladesh and Brazil and have since spread worldwide. In 2014, over 60 countries had some type of CCT program. Some rigorous research has been conducted, in part by J-PAL affiliates, on the impact of CCT programs in the short run. This research has generally found that CCTs are effective in improving incentivised outcomes (See Gertler 2004; Shultz 2004; Baird et al. 2010; Baird et al. 2014). There have also been studies on whether conditions are even necessary, as they are costly to enforce. For example, in Morocco, Benhassine et al. (2015) compared a CCT with a labelled cash transfer (LCT), where the programs were described similarly but conditions were enforced under the CCT and not enforced under the LCT. They found that the outcomes of the CCT and LCT programs were very similar; adding conditionality in this context made almost no difference.

In comparison, in Malawi, Baird, McIntosh and Özler (2011) conducted a randomised evaluation to compare unconditional cash transfers (UCTs) with CCTs, and they found that there were trade-offs with conditionality. Like other researchers, they found that CCTs were effective at incentivising behaviour, but they also found that UCTs were more effective for certain types of individuals. For example, they found that UCTs were more effective in helping girls to delay marriage and childbearing. They surmise that the CCTs successfully incentivised some households to meet the transfer conditions. However, there were other types of households that were so far from being able to complete the conditioned activities that they did not respond to the incentive. Such households might have really benefitted from the cash transfer. In those cases, enforcing conditionality stopped cash transfers to those households, which had a countervailing negative effect on that class of household. While these results confirm that CCTs improve indicators of incentivisation, they also indicate that enforcing conditions may have negative consequences, such as excluding benefits from households that are unable to meet conditions for various reasons.

Medium-term Evidence from Indonesia

Most of the existing evidence on cash transfers looks at the short run, but what happens in the longer run? As noted by Santiago Levy, a pioneer of the Progresa-Oportunidades CCT program in Mexico, 'clearly achieving good health is a cumulative process, temporary investments in nutrition are of little help. The same is true of education: children must be supported year after year' (Levy 2006). This helps to explain why CCT programs typically determine eligibility once and then give transfers to households for many years, as the programs are intended to support children as they grow up. However, it is difficult to study the long-term impact of CCTs. In most cases, the government gave the program to the comparison group after a relatively short period, making it difficult to learn about the long-run impacts of these types of programs. For example, in the well-known case of Progresa-Oportunidades in Mexico, the control group received the program just 18 months after the study began.

In Indonesia, Cahyadi et al. (2018) conducted a medium-run study of the impacts of the PKH. As described above, in the original program design, 736 *kecamatan* were randomised either to receive the PKH transfers or to serve as a comparison group. The two-year results were analysed by Alatas (2011).

It turns out that as the government expanded the PKH after the initial evaluation, it prioritised expanding the program to new provinces and districts in order to ensure that it spread nationwide, rather than prioritising the original comparison group. Thus, without any researcher intervention, by 2013, more than six years after the program had started, all the original treatment locations from the first evaluation were still receiving the PKH transfers, and 60% of the original control group remained untreated. This created a unique opportunity to measure the impact of the CCT program over the longer run.

We found substantial improvements in health and education behaviour in both the short run and the long run. For example, the PKH improved maternal delivery behaviours after two years, and the effects increased in magnitude after six years. While there was no impact on child immunisations after two years, the PKH increased child immunisations after six years. Similarly, for school enrolment of older students (aged 15–17), there were no effects after two years, but there were positive effects after six years. This may be because the original program had not been in place long enough to affect enrolments among older students. However, our study suggests that after six years, the program was able to increase enrolments among older students who had been exposed to the program year after year.

Perhaps most remarkably, after six years there was some evidence of cumulative health impacts, particularly reductions in stunting. Stunting is a major policy challenge in Indonesia and can be reduced only through prolonged attention to weight and nutrition over a child's early life cycle. Our results indicate that the PKH actually reduced stunting after six years. This may be because the children evaluated in the long-term study benefitted from the PKH for a substantial part of their early childhood, where they and their mothers were supported by the program for six years through cash transfers conditional on better maternal health and childhood investments.

Overall, what we found was not that the PKH is radically transforming the economic well-being of households. Rather, the evidence suggests that sustained investments in these households over time can lead to better human capital investments in their children. This in turn could help break the intergenerational cycle of poverty and improve things in the longer run.

We also reassessed Generasi in the same way, almost 10 years after it had started. As with the PKH, when Generasi was expanded, the government prioritised new areas of the country (and, in particular, remoter areas, following the results of the original evaluation). Consequently, the original control group received no special priority in receiving the program, but the program continued in treatment areas, creating an opportunity to evaluate its longer-term impacts.

We found that as Indonesia experienced economic growth, supply-side conditions improved in both the treatment and control areas. Many of the original areas that were initially supply deficient were no longer so 10 years later. As a result, the program, which was most effective in areas with low levels of service delivery at baseline, had substantially muted effects 10 years later (Olken and Sacks 2018). The program was still effective in the areas with the lowest levels of service delivery, particularly in terms of improving child weight checks, immunisations and vitamin A levels. However, in general, the changing health and education environment in Indonesia suggests that the optimal policy mix for a given place may need to change too. The contrasting results—with increasing effects over time for the PKH but decreasing effects over time for Generasi in most locations—suggest the importance of continuing to gather evidence on program effectiveness over the longer term, as the environments in which programs operate evolve.

CONCLUSIONS

Throug these examples, I have described an iterative process between the Indonesian government and both domestic and international scholars in order to identify key policy challenges and generate evidence to improve the design of social programs. During more than a decade-long partnership, we have generated policy lessons for improving the targeting of social protection programs, strengthening transparency and reducing leakages, moving from in-kind transfers to cash transfers, and addressing policy challenges such as stunting. As the challenges have evolved, we have worked closely with the government to answer new policy questions in a rigorous way. As a result, Indonesia has become a leader in generating rigorous, policy-relevant evidence for the rest of the world.

The results of the process described here are therefore useful far beyond Indonesia. Indeed, part of the iterative process is that the topics we prioritised for new studies are those that are not only immediately relevant to Indonesian policymakers but also helpful in contributing to global knowledge. The work in Indonesia has led to renewed academic interest in related targeting questions, which have taken place in multiple countries in sub-Saharan Africa (Dupas et al. 2016; Dizon-Ross, Dupas and Robinson 2017) and elsewhere. As other countries reform their targeting systems, the lessons learned from Indonesia can help inform them as well.

ACKNOWLEDGEMENTS

The results described in this paper draw on my collaborations with many people, including Vivi Alatas, Abhijit Banerjee, Nur Cahyadi, Gabriel Kreindler, Jordan Kyle, Junko Onishi, Rizal Adi Prima, Ririn Purnamasari, Audrey Sacks, Elan Satriawan, Sudarno Sumarto, Ekki Syamsulhakim, Julia Tobias, Matthew Wai-Poi, Susan Wong and particularly Rema Hanna.

I thank our many partners in the Indonesian government, particularly Rudy Prawiradinata, Pungky Sumadi, Maliki and Vivi Yulaswati at Bappenas, Bambang Widianto at TNP2K, and many partners at BPS, the Ministry of Home Affairs and the Ministry of Social Affairs—for their openness to using evidence to inform policy decisions and for their support of the projects described here.

I thank the many staff members at J-PAL Southeast Asia for outstanding work implementing these projects, including Talitha Chairunissa, Masyhur Hilmy, Amri Ilmma, Nurzanty Khadijah, Chaerudin Kodir, Nadira Purdayinta, Héctor Salazar Salame, Ritwiki Sarkar, Prani Sastiono, Freida Siregar, Jurist Tan, Poppy Widyasari and particularly Lina Marliani; Gregorius Kelik Agus Endarso and Yulia Herawati for their help with the Generasi project; and SurveyMeter and Mitra Samya for their partnership on many of these projects. I also thank Yuen Ho, Evelina Smirnitskaya and Claire Walsh for help preparing this paper.

Finally, I thank Sudarno Sumarto for hosting me as a young graduate student at SMERU, studying the OPK/Raskin program in the summer of 2001; Lant Pritchett for introducing me to some of the challenges of social assistance in Indonesia in the fall of 2000; the Henry Luce Foundation for sending me to Indonesia in the first place in 1997–98; and Scott Guggenheim for almost two decades of unwavering support and encouragement.

Funding for the projects described here came from the Australian government (AusAID and DFAT), the Royal Netherlands Embassy, 3ie and the World Bank. All the views expressed in the paper are mine and do not necessarily reflect the views of any of the institutions or individuals acknowledged here.

REFERENCES

Alatas, Vivi. 2011. *Program Keluarga Harapan: Impact Evaluation of Indonesia's Pilot Household Conditional Cash Transfer Program (English).* Washington, DC: World Bank. http://documents.worldbank.org/curated/en/589171468266179965/Program-Keluarga-Harapan-impact-evaluation-of-Indonesias-Pilot-Household-Conditional-Cash-Transfer-Program.

Aker, Jenny C. 2017. 'Comparing Cash and Voucher Transfers in a Humanitarian Context: Evidence from the Democratic Republic of Congo'. *World Bank Economic Review* 31 (1): 44–70.

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 (4): 1206–40.

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.

Baird, Sarah, Ephraim Chirwa, Craig McIntosh and Berk Özler. 2010. 'The Short-term Impacts of a Schooling Conditional Cash Transfer Program on the Sexual Behavior of Young Women'. *Health Economics* 19 (S1): 55–68.

Baird, Sarah, Francisco Ferreira H.G., Berk Özler and Michael Woolcock. 2014. 'Conditional, Unconditional and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes'. *Journal of Development Effectiveness* 6 (1): 1–43.

Baird, Sarah, Craig McIntosh and Berk Özler. 2011. 'Cash or Condition? Evidence from a Cash Transfer Experiment'. *Quarterly Journal of Economics* 126 (4): 1709–53.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert and Christopher Udry. 2015. 'A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries'. *Science* 348 (6236): 1260799.

Banerjee, Abhijit, Rema Hanna, Gabriel Kreindler and Benjamin A. Olken. 2017. 'Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs'. *World Bank Research Observer* 32 (2): 155–84.

Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken and Sudarno Sumarto. 2018. 'Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia'. *Journal of Political Economy* 126 (2): 451–91.

Banerjee, Abhijit, Rema Hanna, Benjamin A. Olken and Sudarno Sumarto'. 2018. 'The (Lack of) Distortionary Effects of Proxy-Means Tests: Results from a Nationwide Experiment in Indonesia'. National Bureau of Economic Research (NBER) Working Paper No. 25362, December.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas and Victor Pouliquen. 2015. 'Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education'. *American Economic Journal: Economic Policy* 7 (3): 86–125.

Blattman, Chris, Julian C. Jamison and Margaret Sheridan. 2017. 'Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia'. *American Economic Review* 107 (4): 1165–206.

Cahyadi, Nur, Rema Hanna, Benjamin A. Olken, Rizal Adi Prima, Elan Satriawan and Ekki Syamsulhakim. 2018. 'Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia'. National Bureau of Economic Research (NBER) Working Paper No. 24670, November.

Cunha, Jesse M., Giacomo De Giorgi and Seema Jayachandran. 2019. 'The Price Effects of Cash versus In-kind Transfers'. *Review of Economic Studies* 86 (1): 240–81.

Dizon-Ross, Rebecca, Pascaline Dupas and Jonathan Robinson. 2017. 'Governance and the Effectiveness of Public Health Subsidies: Evidence from Ghana, Kenya and Uganda'. *Journal of Public Economics* 156: 150–69.

Dupas, Pascaline, Vivian Hoffmann, Michael Kremer and Alix Peterson Zwane. 2016. 'Targeting Health Subsidies through a Nonprice Mechanism: A Randomized Controlled Trial in Kenya'. *Science* (353) 6302: 889–95.

Evans, David K. and Anna Popova. 2017. 'Cash Transfers and Temptation Goods'. *Economic Development and Cultural Change* 65 (2): 189–221.

Gertler, Paul. 2004. 'Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment'. *American Economic Review* 94 (2): 336–41.

Hanna, Rema and Benjamin A. Olken. 2018. 'Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries'. *Journal of Economic Perspectives* 32 (4): 201–26.

Haushofer, Johannes and Jeremy Shapiro. 2016. 'The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya'. *Quarterly Journal of*

Economics 131 (4): 1973–2042. Hidrobo, Melissa, John Hoddinott, Amber Peterman, Amy Margolies and Vanessa Moreira. 2014. 'Cash, Food, or Vouchers? Evidence from a Randomized Experiment in Northern Ecuador'. *Journal of Development Economics* 107: 144–56.

Honorati, Maddalena, Ugo Gentilini and Ruslan G. Yemtsov. 2015. *The State of Social Safety Nets (English).* Washington, DC : World Bank. http://documents.worldbank.org/curated/en/415491467994645020/The-state-of-social-safety-nets-2015.

Jensen, Anders. 2019. 'Employment Structure and the Rise of the Modern Tax System'. National Bureau of Economic Research (NBER) Working Paper no. 25502, January.

Krueger, Alan B. and Bruce D. Meyer. 2002. 'Labor Supply Effects of Social Insurance'. *Handbook of Public Economics* 4: 2327–92.

Levy, Santiago. 2006. *Progress against Poverty: Sustaining Mexico's Progresa-Oportunidades Program*. Washington, DC: Brookings Institution Press.

Moffitt, Robert A. 2002. 'Welfare Programs and Labor Supply'. *Handbook of Public Economics* 4: 2393–430.

Olken, Benjamin A. 2006. 'Corruption and the Costs of Redistribution: Micro Evidence from Indonesia'. *Journal of Public Economics* 90 (4–5): 853–70.

Olken, Benjamin A., Junko Onishi and Susan Wong. 2014. 'Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia'. *American Economic Journal: Applied Economics* 6 (4): 1–34.

Olken, Benjamin and Audrey Sacks. 2018. *Indonesia - Long-term Impact Evaluation of Generasi (English)*. Washington, DC: World Bank.

Shultz, T. Paul. 2004. 'School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program'. *Journal of Development Economics* 74 (1): 199–250.

TNP2K (National Team for the Acceleration of Poverty Reduction). 2019. 'Alokasi Anggaran Program Perlindungan Social' [Budget allocation for the social protection program]. PowerPoint presentation.

Triyana, Margaret. 2016. 'Do Health Care Providers Respond to Demand-side Incentives? Evidence from Indonesia'. *American Economic Journal: Economic Policy* 8 (4): 255–88.

World Bank. 2017. *Closing the Gap: The State of Social Safety Nets 2017.* Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/26655.