CHAPTER 7

Designing Social Protection Programs: Using Theory and Experimentation to Understand How to Help Combat Poverty

R. Hanna*[,1], D. Karlan[§]

*Harvard University, Cambridge, MA, United States
[§]Yale University, New Haven, CT, United States
[1]Corresponding author: E-mail: Rema_Hanna@hks.harvard.edu

Contents

Handbook of Economic Field Experiments, Volume 2
ISSN 2214-658X, http://dx.doi.org/10.1016/bs.hefe.2016.07.002

Abstract

"Antipoverty" programs come in many varieties, ranging from multifaceted, complex programs to more simple cash transfers. Articulating and understanding the root problem motivating government and nongovernmental organization intervention are critical for choosing among many antipoverty policies or combinations thereof. Policies should differ depending on whether the underlying problem is about uninsured shocks, liquidity constraints, information failures, or some combination of all of the above. Experimental designs and thoughtful data collection can help diagnose the root problems better, thus providing better predictions for what antipoverty programs to employ in specific conditions and contexts. However, the more complex theories are likewise more challenging to test, requiring larger samples, and often more nuanced experimental designs, as well as detailed data on many aspects of household and community behavior and outcomes. We provide guidance on these design and testing issues for social protection programs, from how to target programs, to who should implement the program, and to whether and what conditions to require for program participation. In short, careful experimentation-designed testing can help provide a stronger conceptual understanding of why programs do or not work, thereby allowing one to ultimately make stronger policy prescriptions that further the goal of poverty reduction.

Keywords

Social protection; Development; Antipoverty

JEL Codes

O10; O12; H53

1. INTRODUCTION

In low-income countries, more than one billion individuals are enrolled in at least one safety net program (Gentilini et al., 2014).[1] These programs come in various forms and sizes. Some aim to simply supplement consumption in hard times. Other newer, more nuanced, social protection programs aim to address the underlying market failures that may have contributed to a household's persistent state of poverty in the first place, driven by a belief that directly addressing these failures may help families break out of a poverty trap. The ultimate choice of program—or combination therein—that countries choose to implement will depend greatly on their social goals, institutional capabilities, and resources. However, even within each broad category of program, the specific design choices made and methods of implementation may affect whether these programs actually achieve their stated goals.

To start thinking about how to design—and then test the impacts of—safety net programs, we begin by classifying them into four main categories based on the underlying motivation for intervening. The first and simplest category is comprised of programs designed for redistributive purposes, e.g., recognizing that the marginal utility of

[1] Throughout this chapter we refer to "social protection" and "safety net" programs as one and the same.

consumption is higher for the poor than for the rich, and thus transfers are socially optimal from a utilitarian perspective (assuming naturally that the taxation process does not create considerable deadweight losses). While these programs can differ in actual design, they share the common feature of first identifying the neediest families along a particular metric and then providing them with cash transfers. Of course, these programs could still have long-run growth impacts, e.g., if they are large enough to provide sufficient capital to start new agricultural, migratory, or business activities (e.g., Banerjee and Newman, 1993; McKenzie and Woodruff, 2006) or if they persuade risk averse households to invest in riskier but more profitable endeavors (Chetty and Looney, 2006) and so forth.[2] But, the primary goal of these types of programs is simply to limit poverty and hunger by ensuring that households attain a minimum living standard.

Second, a missing insurance market may motivate a social protection program. The poor faces many risks, such as unexpected health costs, agricultural damage, or job losses. Without fully functioning financial markets, households may be unable to borrow to smooth consumption. Informal credit and insurance markets provide another avenue to do so, but they often underperform and have become even less effective as countries grow and urbanization breaks down traditional social networks (Coady, 2004). Furthermore, even if households smooth consumption in the face of income shocks, they may be doing so in the short run by making long-term sacrifices, such as pulling children out of school. The insurance market may exist but needs a subsidy to increase quantity demanded, or it may not exist at all and the safety net program directly provides protection in bad times. Social protection from natural disasters is an extreme example of the insurance motivation; however, economic theory and empirical work have had less to say about the structure of such policies. Finally, unemployment and health insurance programs also fit into this often-(at least partially) missing insurance market category.

Third, there may be behavioral or household-bargaining constraints that influence the choices of low-income households and perpetuate poverty. For example, difficulty in resisting immediate temptations can lead to undersaving and thus underinvestment in lumpy goods. Intrahousehold-bargaining issues can lead to suboptimal outcomes for the underpowered, typically women and thus children. Many transfer programs aim to account for these behavioral factors, by providing in-kind transfers (such as food) rather than cash to prevent temptation or by providing transfers to women rather than men to increase female-bargaining powering in the household. Importantly, many programs—conditional cash or in-kind programs—even directly condition assistance to poor families

[2] Notable examples that have found changes increase in investment as a result of cash transfer programs include Covarrubias et al. (2012) who find that Malawi's Social Cash Transfer led to an increase in agricultural investments; and Gertler et al. (2012) who find that Mexico's CCT (Progresa) led to higher levels of agricultural income, as households partially invested a portion of the cash in productive assets.

on behaviors that society would like to address, e.g., children attending school and receiving vaccinations or other preventive health measures. Finally, "workfare" programs also share this aim, by providing transfers conditional on labor force training and/or participation.

Fourth, and finally, there may be market failures preventing asset accumulation. We focus on two domains where asset accumulation could be important for social protection. First, short- and medium-run productive asset building could increase household income, thus allowing individuals to no longer need consumption transfers from government. Rather than cash transfers, productive asset transfers as well as training, coaching, and informational programs may be critical to motivate and improve investment choices to make such a goal attainable. Second, long-term financial asset building may be difficult if savings markets are missing, or if behavioral and household constraints discussed earlier bind. The current savings market infrastructure does not facilitate pension savings for individuals, and building such markets could be critical for improving consumption for the elderly in developing countries, just as in developed countries.

Naturally many programs can and often do target multiple issues at once, for two reasons. First, individuals may face multiple market failures, thus motivating a more complex program. For example, recently multifaceted programs have arisen that explicitly aim to "graduate" households out of extreme poverty by providing households with different inputs, including working capital, assets, and jobs training. These programs aim to increase household earnings capacity by concurrently relieving a number of different barriers to economic growth. Second, low-income households even within a particular setting may face different market failures and thus the optimal policy may not be the same for all. This difference across households may motivate targeting different aspects of a program to different people or designing programs that manage to address different issues for different people.

In short, a wide range of tools are available to policy-makers in the goal of poverty alleviation. This variety naturally causes us to question: what are the right programs, who should they be targeted to, and how do we know that they are working? Randomized evaluations can answer this question by providing clear answers to whether or not a program "works." However, importantly, a randomized evaluation can go even further, providing insights into why it works, i.e., the underlying mechanisms that drive the observed treatment effects. By doing so, the evaluation can offer greater insight into whether a similar program would work elsewhere or how the program should change as circumstances change even within the same context. To generate these insights, a well-designed evaluation must pay careful attention to the theory behind the different forms of social protection programs, consider multiple treatment variations to isolate theoretical channels, and be creative in data collection.

We will first discuss issues to consider in testing programs that fall under each of the four categories we lay about earlier, weaving in examples and knowledge from the current state of the literature.[3] We then address issues pertaining to implementation, as success not only relies on whether the proposed program theoretically can achieve the social goal it was designed to address but also relies on how it is implemented in practice. Finally, we offer advice on further research needed in this space, including a better understanding of both the general equilibrium and long-run impacts of these programs.

2. REDISTRIBUTIVE PROGRAMS

Motivations for redistribution, absent specific market failures, inevitably come from social preferences for lower levels of inequality or from philosophical tenets such as utilitarianism. In the simplest utilitarian form, elegantly put forward by Peter Singer (1997), redistribution is motivated by an awareness of the stark trade-offs between more of one's own consumption (for the wealthy) versus more consumption for the poor. Philosophical motivations abound, naturally: for example, John Rawls' *A Theory of Justice* (1971) argues for a veil of ignorance in which one makes moral and policy judgments without knowledge of one's own position in society. This hypothetical construct leads to arguments for redistribution to the worst-off members of society.

With even minimal weighting on other people's utility in one's own utility, some redistribution typically becomes the individually optimal policy for all. This naturally motivates why an individual may redistribute their wealth to others (i.e., through charitable actions), but not necessarily why a government, through taxation, may effectively mandate such redistribution. However, there are many reasons why a government may mandate such redistribution, rather than leave it to the voluntary actions of individuals. First, transaction costs may be high for individuals to target the poorest effectively. Second, it may be costly for wealthy individuals to transfer wealth to the poorest. Third, high levels of inequality can cause social strife, or even violent uprising, and collective action problems make it difficult for a purely voluntary system to generate sufficient redistribution to achieve the social optimal. Fourth, behavioral theories could explain why individuals underredistribute if left to their own device but do have a stated preference for more redistribution. This is akin to models of quasi-hyperbolic preferences or the dual self (Fudenberg and Levine, 2006) applied to redistribution: if one is maximizing the welfare of the more deliberative and sharing self (as opposed to the impulsive and selfish self), then one may want a commitment device to help stay in line with their more deliberative preferences. A government program for redistribution thus becomes such a device. Fifth,

[3] Summarizing the broad and sizable literature on poverty alleviation programs poses a unique set of challenges. We have tried to focus on key ideas and use the literature to help provide examples when possible, as well as to give insight into where open questions persist. We have tried to cover the important papers in the literature but cannot cover all of them while keeping this chapter a reasonable length. We apologize in advance to those whose papers that we do not cover in detail.

individuals may be misinformed about the current level of inequality and poverty, and their relative wealth, whereas policy-makers may be more informed. In the United States, for example, recent work shows that most people vastly underestimate the level of inequality in the United States and state a preference for more equality, even though they are simultaneously opposed to more taxes. This finding demonstrates a clear information gap (Norton and Ariely, 2011; Kuziemko et al., 2015).

In short, there are a number of rationales for governments to engage in redistributive activities. A key empirical question is how to best design these programs and test if they accomplish their goals. In doing so, the first aspect to consider is how to identify the poor to direct resources toward them ("targeting"). In Section 2.1, we first discuss the potential strengths and weaknesses of common targeting methodologies, as well as describe the key factors that one must consider when planning randomized control trials (RCTs) on targeting methodologies. In Section 2.2, we then discuss how to redistribute: for example, how big should the transfers be? How long should the transfers last? In the process of this discussion, we recount what we know and do not know from the current experimental literature. We also provide a guide for the design of RCTs to evaluate the types of social protection programs outlined earlier.

2.1 How to target the poor?

In high-income countries, targeting is often achieved by means testing: households bring a proof of income or unemployment to a benefits office, or they receive transfers through tax systems, such as through the United States Earned Income Tax Credit. However, in low-income countries, a lack of formal labor markets with a paper trail of income and employment status, coupled with underdeveloped tax systems, results in limited data to verify income.

To fill this data gap, low-income country's governments can conduct income or consumption censuses. However, such censuses also present their own set of challenges, as anyone who has ever tried to conduct a survey module to elicit these kinds of data can attest: the modules take an inordinate amount of time and require certain skills, since one needs to map out all the different components of income (e.g., farming your own land one day, casual labor the next) or consumption (e.g., items purchased, the crops that one grows). Plus, without a formal mechanism with which to cross-check data, there is often nothing stopping surveyed populations from lying if they know that a cash prize is attached to their answers.

As such, low-income countries tend to develop alternative methods with which to identify the poor. The method chosen will depend on the priorities of the government: for example, Is the aim to target based on a particular poverty line?[4] Is it preferable to target

[4] Note that in addition to targeting poverty status, some programs also target particular demographic characteristics, such as whether a woman is pregnant or has children. As targeting on these characteristics tends to be easier—due to their verifiability—we will not discuss them in detail here for conciseness.

the poor based on income, consumption, or some other metric of poverty? How spatially dispersed are the poor? It will also depend on the institutions in place and context: how good is the implementing agency's ability to conduct surveys? How responsive are local leaders to citizens? We first outline key categories of targeting methodologies and then discuss how experimental methods can help distinguish between varying features of these methods.

Finally, note that we focus on targeting *poor* households in this section, given that the goal of redistributive programs is to provide impoverished households with a basic standard of living. Transfer programs that hope to enact longer run changes may target differently, depending again on their goals. For example, Barrera-Osorio and Filmer (2013) compare the effectiveness of scholarships when they are targeted to the poor versus when they are merit-based: while both increase school enrollment, only merit scholarships increase test scores. Thus, the method that you choose would depend upon whether you would like to redistribute scholarships to the poor, or to redistribute scholarships to those who will have the highest marginal return from them given a metric of test scores. We will revisit the idea of differences in targeting methods and goals later, when we discuss programs that aim to change long-run outcomes.

2.1.1 Targeting methods

While there are many variations in practice, there are four primary categories that encompass most targeting methodologies:

- Geographic: If the poor are concentrated in particular villages, districts, or regions, giving everyone within those areas access to social protection may be an effective method to transfer resources to the poor (see, for example, Baker and Grosh, 1994; Elbers et al., 2007). Moreover, this form of targeting may be particularly attractive when the institutional capacity needed to collect individual information is low, as only aggregate information is needed, e.g., poverty mappings, rainfall data, etc. Note, however, that this method may also be politically sensitive as it disburses benefits to some areas but not others.
- Proxy-means testing (PMT): In this method, the government collects demographic and asset data from households and uses these data to predict or "proxy" income or consumption.[5] Sometimes this method involves a quick-and-dirty poverty score card with just a few questions. Other times, a longer, more detailed asset and

[5] This is typically done with a nationally representative data set that includes the variable on which to target (e.g., income), as well as numerous demographic and asset variables. Next, income is regressed on different household characteristics, looping through different combinations and permutations of the variables, until the set of characteristics that best predicts income is identified you find (often regional fixed effects or regional variables are also included for better precision). After conducting a census to obtain the chosen household characteristics, it is then possible to compute predicted income for each household using the formula. Households below a chosen cutoff of predicted income would thus be eligible for the program.

demographic survey is conducted. However, in either case, the key is to choose variables that are simple to collect, relatively easy to verify (e.g., whether the household has a dirt or concrete floor, or if they have telephone line), and that are less likely to be distortionary (e.g., school enrollment may predict poverty, but we may not want to incentivize households to keep their kids out of school). Households who pass the PMT—i.e., are below a certain poverty line—are then automatically enrolled. The effectiveness of the PMT will depend upon different factors: the formula's predictive power, the quality of survey team, etc.

- Self-targeting: Self-targeted programs are those in which everyone is allowed to apply, but in which some sort of "barrier" is put into place to reduce the probability that rich households try to access the program. Theoretically, there are different barriers that can generate this form of selection (Nichols and Zeckhauser, 1982), ranging from time costs to apply, means testing on arrival (Alatas et al., 2016), and work requirements (Besley and Coate, 1992; Ravallion, 1991).[6] When done right, this can effectively screen out the rich (see, for example, Alatas et al., 2016; Christian, 2014). Given that households may fall into and out of poverty, these programs also have the potential advantage of flexibility, allowing households to access the programs when they have a bad shock. However, these programs also run a substantial risk: it may be hard for governments to initially predict, in advance, how many people will actually enroll or participate, providing additional challenges to budgeting and implementation.
- Community-based methods: In this method, community members choose who are needy in their locality. Theoretically, this method could not only bring in better local information on who is poor but it could also incorporate the community's perceptions as to what determines poverty in their location (Seabright, 1996). A potential benefit is that this may make the program more politically popular, as people may feel that the list is more in-line with their vision of who is needy or deserving. A key worry is that by allowing for local discretion in choosing program recipients, elites may possibly capture the process (Bardhan and Mookherjee, 2000). Moreover, another potential downside is that while this method elicits better data on relative poverty within an area, it does not provide information across villages. For example, Alatas et al. (2012) show that the difference between the PMT's and community's ability to target based on consumption nearly doubles when the PMT is allowed to use its cross-village information.

Ultimately, the method and design will depend on a number of factors: the goals, context, institutional capacity, targeting budget, etc. For example, Alatas et al. (2012) show that the choice of community methods versus PMT can depend on whether one

[6] For certain types of in-kind goods—subsidized health products, insurance products—the price charged may also be used to select a particularly type of person who may value the particular product (see, for example, Cohen and Dupas, 2010; Beaman et al., 2014). We discuss this in more detail in Section 4, when we discuss programs that are geared at changing longer-run income or behavior.

wants to specifically target a hard measure of poverty (e.g., income) or a soft one (e.g., perceptions). Similarly, choosing the right design of the PMT—e.g., the number of questions that goes into the formula—will also depend on the institutional capacity to administer the PMT. Moreover, Beath et al. (2013) show that whether aid allocations reach the neediest or not can depend on the type of community institutions that participate in targeting.

In practice, while there are distinct categories of methods, governments often mix and match the methods depending on circumstance. For example, they often try to save money by conducting a PMT on a selected sample of people who are likely poor, rather than conducting a full PMT census. For example, in its earlier form, Mexico's Progresa program conducted a PMT to determine eligibility only in the areas that were chosen as likely poor based on geographic targeting (Schultz, 2004). Similarly, Indonesia's Data Collection on Social Protection Programme (PPLS) uses community-based methods to help determine the list of households that will receive the PMT survey (Alatas et al., 2012). Kenya's Cash Transfer for Orphans and Vulnerable Children actually uses three methods: first, geographic targeting is used to determine locations, then community targeting is used in the selected areas to determine a list of households, and finally, those households are given a PMT to determine actual eligibility (The Kenya CT-OVC Evaluation Team, 2012).

2.1.2 Experimentally testing between targeting methods

Are experimental methods important for testing across targeting approaches? Not necessarily. For example, one simple research design would just be to try out two different methodologies in the same areas and then compare the income levels of those selected under both methods.

While this research strategy may prove attractive in some ways, it may miss out on the nuances of targeting that may ultimately be quite important in understanding the relative effectiveness of the differing methods. First, many of these targeting methods require considerable effort on the part of citizens and the program staff.[7] Simulating conditions to be as real as possible, with at least a small amount of cash on the line for the households that are to be selected, is important to understand how the methods would actually work in practice.[8] In terms of citizens, people might behave differently during the process if they do not believe that the targeting list will actually be used to distribute resources.

[7] One obvious exception to this is geographic targeting, which does not rely on field operations, such as household and community leader interactions. If a different method was conducted for the actual program, then simulating geographic targeting over the same areas based on administrative data can provide an accurate comparison of the type of person selected under both methods. However, by not conducting geographic targeting in practice, it will not be feasible to test program outcomes other than just who is selected (e.g., political acceptability, leakages) that result from using different targeting methods.

[8] You could, for example, try out two methods in the same area and offer a transfer to everyone who is selected by either of the different methods, but then staff or villages may coordinate so that different people are on each list. And, it runs the risk of confusing people, so that they do not take the exercise very seriously.

For example, they may not exert any effort in discussing and ranking households in community targeting, or individuals under self-targeting may just not bother to show up even though they would, if there were real cash involved. Moreover, people may give more truthful answers in the PMT, or not claim a greater poverty status than their reality at a community meeting, if they believe that their answers do not have any consequence. In terms of staff, the results may noticeably differ if you test out methods without stakes with a highly trained set of staff than if trying out the methods with the typical type of staff that would be hired and trained (Alatas et al., 2012).

Second, the choice of targeting method may affect both the program and household outcomes. For example, the targeting method chosen may help determine the ultimate satisfaction of the program, which in turn may affect program acceptance and the government's ability to implement the program. Targeted programs can be controversial, with some households receiving a transfer, while their neighbors do not. If citizens believe that a certain method is unfair and that it would produce a flawed list, they may be less likely to support the overall social protection program and potentially block the distribution of benefits.[9] Moreover, if people believe that the wrong individuals were chosen due to the fact that a certain method was used, it may possibly lead to distortions in how informal insurance or lending operates within a community.

Finally, who is chosen may be different than who actually receives the transfer (Alatas et al., 2013) and this may also vary by the targeting method that was employed. For example, suppose that the PMT better selected the poor than a community method. But that the PMT had less legitimacy than the community method, so that village leaders did not adhere to the PMT in practice when distributing the transfers, but they did adhere to the community list. In this case, simply simulating both methods in the same area to elicit a beneficiary list would possibly wrongly suggest that one should select the PMT since you would only be able to study who was chosen but not be able to measure who ultimately would get the benefits.

In short, for all of these reasons, we would want to randomize the targeting methods to different areas to see how the method affects who is chosen and who ultimately receives the transfers. The optimal design for a field experiment in the domain of targeting will depend not only on a number of factors, including which method is being studied, but also what the particular context looks like. However, there are a number of key questions to keep in mind regardless of the given design.

The first question is whether or not it is necessary to have a control group, in the traditional sense. RCTs generally compare the outcomes from a treatment group that

[9] For example, Alatas et al. (2012) show that community targeting led to much higher levels of satisfaction than the PMT, with village leaders feeling less comfortable making the transfers under the guise of public scrutiny when the PMT had been used. They provide suggestive evidence that this difference is due to the perceptions of the methods, and not the ultimate lists that the different methods produced.

receives an intervention with those of a control group that does not. In this case, since the outcome of interest is who is selected under different targeting methods within the same program, it may be viable to simply have multiple treatment groups where each is randomly assigned a different targeting method, but everyone receives the program.

Second, at what level should the randomization take place? Randomizing at the individual level offers the most statistical power, but in this case, the targeting treatments often involve some sort of group participation (e.g., community, self-targeting) or group data (e.g., geographic). Moreover, even with a PMT, where it is possible to vary how the survey is conducted across individuals, a question of interest might also be whether the program administrators change how they respond to the targeting list in their area based on which method generated the list. Thus, in most cases, it is appropriate to randomize across a sensible geographic unit; as such, power calculations to determine sample size should account for the group structure of the data.

Third, is a baseline survey needed? With most experiments, the answer is not necessarily: the analysis will consist of comparing outcomes of those in the treatment and the control group. A baseline might help for power if the outcome measures within a person are highly correlated across time and it may allow you to test for the heterogeneous treatment effects by various baseline characteristics (Duflo et al., 2006). But, it is not necessary per se in a typical randomized experiment. In this case, a baseline is essential: a key outcome of any targeting experiment will be the baseline income or consumption levels—prior to the targeting—of those who are actually chosen.

Fourth, what kinds of data should be collected? The exact variables would, of course, depend on what methods are being tested and what outcomes are expected. But, typically, in the baseline, it is essential that data be collected on the variables being targeted on (e.g., income, consumption, etc.), so that inclusion and exclusion errors can be computed.[10] To measure distortions in who is chosen along certain dimensions, it may be worthwhile to collect baseline data on political affiliations and relations to political leaders. In the end line surveys, valuable data to collect could include who actually received the program, satisfaction levels, and metrics on general program functioning.

Finally, will the experiment just aim to measure the reduced form effect or will it also attempt to understand why it is working (or not)? If the only relevant question is comparing method one versus method two (i.e., the reduced form difference of the two programs), only two treatment arms are needed. But we know that the effectiveness of a method may vary based on its own design, and therefore, it may be worthwhile to learn more about a method's effectiveness if certain details of the implementation are varied. Here, theory can help guide the appropriate subrandomizations: For example,

[10] Alatas et al. (2012) also ask villagers to rank one another to gain the "average" person's belief about another household's poverty status in a village, as well as ask household to assess their own poverty status.

Alatas et al. (2016) experimentally compare the outcomes of a PMT with a self-targeting mechanism within Indonesia's conditional cash transfer (CCT) program. Importantly, they experimentally vary the distance of the application site under self-targeting to generate exogenous variation in the cost of the application "barrier." They then use this variation to estimate a model of the decision to apply for the program and simulate self-targeting outcomes under different levels and types of application costs.

2.2 Evaluating the impacts of redistributive programs

Once the poor have been identified, the next task at hand is to think about whether the redistributive programs are achieving their goals in practice. The simplest redistributive programs are those that entitle the identified poor to some form of cash stipend to provide a certain standard of living and are *unconditional* on any behaviors (an unconditional cash transfer, or UCT). The Chinese Di-Bao Program is the largest UCT in the developing world, reaching 78 million households (see Chen et al. (2006) for a description of the program). Other prominent examples include South Africa's Child Support Grant, India's National Old Age Pension Scheme, and Kenya's Hunger Safety Net Program.

Evaluating these programs usually follows a typical design. First, poor households are targeted in the sample area. Then, potential beneficiaries are randomly assigned to the treatment group ("receives transfers") and the control group ("does not receive transfers") to assess the program impacts. Examples of unconditional cash programs that have been evaluated in this fashion include the Zambia Child grant program (Jessee et al., 2013), the Kenya Hunger Safety Net program (Merttens et al., 2013), Kenya's Cash Transfer for Orphans and Vulnerable Children (see, for example, The Kenya CT-OVC Evaluation Team (2012) and Covarrubias et al. (2012), on the Malawi Social Cash Transfer Scheme).[11]

Even though the overall evaluation strategy is relatively straightforward, there are still a number of decisions to be made—concerning the level of randomization, the type of data collected, the timing of data collection, etc.,—that are important in assessing impacts.

The level and form of randomization: Transfers are generally provided to an individual or household, and so it is tempting to randomize at the individual level to maximize statistical power. For example, Schady et al. (2008) do exactly this. However, Angelucci and de Giorgi (2009), among others, show that there may be spillovers of transfers from

[11] There were two RCTs conducted on *Bono de Desarrollo Humano*—one on child health (Paxson and Schady, 2010; Fernald and Hidrobo, 2011) and one on education (Schady et al., 2008; Edmonds and Schady, 2012). The program was initially supposed to be conducted as a conditional cash transfer program and some announcements were made to this effect (Paxson and Schady (2010) had multiple treatment arms to test between pure cash and cash with conditions), but the conditions were never enforced. Given that just the framing of the program as a CCT could still have impacts, we do not include this in our discussion of UCTs but instead discuss it later.

eligible to ineligible households within a village.[12] This implies that the randomization likely needs to be done at a higher level, at the village level or subdistrict level depending upon what types of spillovers one might expect.[13]

If randomization is at a group level, it is vital to have enough "units" to randomize over or it will be challenging to measure impacts. For example, the Kenya's Cash Transfer for Orphans and Vulnerable Children (The Kenya CT-OVC Evaluation Team, 2012) had only 28 units of randomization, which might account for their difficulty in detecting any program impacts; the Malawi Social Cash Transfer Scheme had only 8 clusters and did not fully account for the grouped nature of the data in the analysis (Covarrubias et al., 2012).[14] Thus, one should determine in advance what the desired size of treatment effects is (i.e., large enough for the program to be cost-effective, etc.) and use this to assess the statistical power of the proposed design.

Importantly, it is also key to identify the beneficiaries in the control group, not just the treatment group. Suppose a randomization determined which villages obtained a UCT and which were in the control group. In the UCT villages, targeting would have first been conducted to choose the beneficiaries within the village. However, unless a similar targeting strategy was also conducted in advance for the control group, it would be difficult to know who were the hypothetical beneficiaries in the control group (see, for example, Covarrubias et al., 2012). In this case, it would be possible to estimate the impact on the entire village as a whole but not easy to estimate the impact on just the beneficiaries.[15] Thus, it is essential, when possible, to use the same targeting methods in the treatment and control group, even if the control group is not receiving the program at the time of the study.

Data collection: What data should be collected and when? If the goal is simply redistributive, we may simply care whether poor households were actually identified and whether or not they were receiving their entitlements. In many developing countries, due to weaker institutional structures, corruption, or imperfect information on their entitlements, households do not receive the full transfer. So, an important set of survey questions should be focused on whether households actually received the transfer, whether they received the full transfer, whether the village elites held them up for a portion of

[12] As we discuss later, Angelucci and de Giorgi (2009) evaluate a conditional cash transfer program, but the basic ideas on spillovers hold for UCTs as well.

[13] As we discuss later, one might even want to design the study to capture different types of spillovers and general equilibrium effects.

[14] In the analysis of these cases, it would be necessary to adjust standard errors to account for both the grouped nature of the randomization and the small number of groups in keeping with the procedure outlined by Cameron and Miller (2010).

[15] Of course, this is not a problem within the village if the program is geographically targeted and everyone in all sample villages is eligible. However, this may also generate spillovers over larger geographic regions that one may want to be aware of, e.g., if everyone in a district of province receives additional income, would this increase demand for food, raising food process?.

their transfer, and so forth. Next, one may want to administer a household income or consumption model to measure household's income status.

If we simply care about redistribution, we would not necessarily care about how households spend the transfers—we would just care about whether they receive the money. Thus, if motivated by such a philosophy, one may argue to only measure the transfers and not bother examining what happens with the money transferred. But for many reasons, researchers do collect more data. For example, one political rationale of many redistributive programs is to provide a basic standard of living for children, so one may want to collect measures on child health, nutrition, and education. There are a number of issues to consider in doing this type of larger data collection, but we will revisit it later when we discuss broader issues surrounding behavioral change for program beneficiaries.

Next, a key question is whether or not a baseline survey is needed. In assessing the impact of a transfer programs, a baseline survey is not necessarily needed, since treatment was assigned randomly, meaning that a postintervention of the two groups is an unbiased estimate of the true impact. However, there are two key exceptions to this. First, if we want to assess whether the poor were indeed targeted and whether the poorest of the poor were able to receive their entitlements, then we would need baseline consumption or income data. Second, if particularly vulnerable subgroups are of importance (e.g., the very poor, families with children who are less likely to attend school, etc.), one may want to collect data on these groups—if administrative data of this sort is unavailable—to be able to stratify the randomization.

Finally, when do you conduct the follow-up surveys? Again, it depends a bit on the research aims. A follow-up survey should generally be conducted within the duration of the program to understand whether households are receiving the transfer. However, as we discuss later, if the longer-run impacts of transfer programs are of interest, one may also want to do additional surveys sometime after a household is no longer enrolled in the program.

3. MISSING INSURANCE MARKETS

The poor face many risks, such as unexpected health costs, agricultural damage, or job losses. Without fully functioning financial markets, households may be unable to borrow to smooth consumption. Furthermore, even if households do smooth consumption in the face of shocks, they may be doing so in the short run by making longer-term sacrifices, such as pulling children out of school or not investing in health. At the extreme, natural disasters—e.g., earthquakes, flooding, famine—may cause group shocks that further limit the functioning ability of informal insurance mechanisms within a region.

Thus, an important role of social protection may be to help alleviate the impact of such shocks by providing social insurance of various forms. Two prominent forms of insurance are agricultural insurance and health insurance, but we will not cover them here since they are discussed elsewhere in this handbook. Rather we will briefly discuss two other forms of insurance: disaster relief and unemployment insurance (UI).

Disaster relief is an important form of social insurance. While some quasi-experimental work has been done to study both the consequences of a disaster, and the distribution of aid, to our knowledge, there has been less experimental evidence in evaluating social protection in humanitarian settings. Notable exceptions include de Mel et al. (2008), who studied the effect of cash transfers to microenterprises in posttsunami Sri Lanka, as well as Aker (2014) and Hidrobo et al. (2014) who examined cash versus other types of transfer programs in informal refugee camps in the Democratic Republic of Congo and Northern Ecuador, respectively. Part of the reason for the lack of experimental evidence comes from ethical concerns with a need to provide immediate assistance trumping research and evaluation. Related is a logistical challenge: the chaos and highly transient nature of refugee camps can make randomization particularly challenging to implement. However, given the about 60 million refugees and internally displaced people worldwide (UNHCR, 2014), understanding how safety net programs can be better tailored to provide assistance in humanitarian settings has been an increasingly urgent need, particularly as many refugee camps persist for long after the immediacy of the disaster.

A second important form of insurance is UI, the provision of temporary assistance to those out of work. This form of social protection has been traditionally missing from poor countries, as the data-poor environments preclude easy identification of those who are working in full-time positions to then determine who is temporarily out of work. But, it is becoming more common in relatively more developed, low-income countries with formal labor markets (e.g., Brazil, Egypt). While there have been numerous experiments that have tried to understand aspects of UI in developed countries, to our knowledge, there is little experimental evidence on UI in lower to middle income settings.[16] One notable exception is Mickelwright and Nagy (2010), which randomized unemployment beneficiaries in Hungary to varying levels of monitoring (i.e., visiting the employment office every 3 months with no job search questions versus visiting every 3 weeks to answer questions on job search behavior). However, as UI spreads and becomes a more important component of social protection in developing countries, empirical evidence will be needed to understand its impacts on labor markets and household outcomes. Empirical evidence will also be needed to understand how to better design these programs—addressing how to best verify employment status, how to best distribute benefits, and how to ensure that the programs provide incentives to find work.

[16] For example, Meyer (1995) summarizes a number of early experiments on UI in the United States, focusing on four cash bonus experiments and six job search experiments. More recent examples include Van den Berg and Van der Klaauw (2006), who explored the effect of additional monitoring and counseling on UI recipients in the Netherlands; Grenier and Pattanayak (2011), who measured the impact of legal assistance on UI in the United States.

4. BEHAVIORAL CONSTRAINTS

It is often argued that behavioral or household-bargaining constraints influence the choices of low-income households, thereby perpetuating poverty. Regardless of whether or not these behavioral constraints are present, safety net programs are often designed to correct these types of constraints and encourage socially optimal behavior.

There are two key forms that behaviorally focused social protection programs usually take. The first is providing in-kind transfers rather than cash, under assumptions such as people will be too tempted to spend money on "bad" goods (such as tobacco and alcohol) or that household-bargaining constraints would imply that cash transfers would cause socially undesirable outcomes, such as an underinvestment in children or an exit from the labor market.

The second is to directly condition the receipt of transfers to households on a household's compliance with certain long-term investments, such as the children attending school or visiting health clinics. These CCT programs have become increasingly common in Latin America and have begun to spread to other parts of the world, existing in more than 52 developing countries as of 2013 (Fiszbein and Schady, 2009; Saavedra and Garcia, 2012; Gentilini et al., 2014). And, finally, note that some programs do a combination of both in-kind and conditions, with 130 countries doing some sort of conditional, in-kind program.

Careful experimentation that helps test the underlying rationale for these programs can help determine if behavioral constraints exist, and if so, what form they take. Thus, they can provide insights into whether these kind of constraints on behavior actually improve welfare or simply make redistributive programs more expensive.

4.1 Providing in-kind transfers to constrain spending choices

4.1.1 The rationale for in-kind programs

Unconditional, in-kind redistributive programs typically provide free or highly subsidized goods to program recipients. They can entail a direct provision of the goods, such as a food or fuel transfer program; the subsidy of a product distributed through local governments, NGOs, or designated shops; or a system of vouchers that are constrained to particular types of goods, such as a food stamp program.[17] Gentilini et al. (2014) note that 89 low-income countries have unconditional, in-kind transfer programs, with examples

[17] It is important to note that there are other types of in-kind transfers, such as ones that provide free health products (Cohen and Dupas, 2010; Dupas et al., 2013; Ma et al., 2013; Glewwe et al., 2014), prizes or scholarships for school (Berry, 2014; Kremer et al., 2009), school meals programs (Kazianga et al., 2012; Vermeersch and Kremer, 2004), and public housing (Kling et al., 2004). These programs may also have different aims, such as solving the externality issue in health product take-up. As they are discussed in other chapters in this handbook, we refrain from discussing them here.

including Indonesia's Rice Subsidy Program ("Raskin") and the Public Distribution Systems in both Bangladesh and India.

There is much debate about whether transfers should be in-kind. If you ask a typical economist, most will favor cash programs, under the idea that households will maximize utility if they have choice over what they purchase, rather than receiving a good of equal monetary worth that they may not value as much. However, there are a number of arguments proposed in favor of in-kind subsidies (see Currie and Gahvari (2008) for an excellent review), which may explain their general persistence worldwide.

The most cited explanation is that of paternalism: People often have an image of the lazy, out-of-work husband co-opting the family cash and wasting it on alcohol, tobacco, and other forms of entertainment, rather than making spending decisions that can improve the family's living situation or investing in children. Thus, the argument follows that an in-kind subsidy could reduce the husband's ability to do so, forcing redistribution within the household to the women and children who typically have less household-bargaining power. However, others argue that if the in-kind subsidy is inframarginal—or if it is easy to resell goods—then in-kind subsidies will not alter the household's consumption bundle, and it is simply a more costly mechanism to redistribute to the poor than cash. The experimental evidence thus far suggests that cash programs do not generate more spending on temptation goods, such as alcohol and tobacco. Nonetheless, in-kind programs are often "sold" as targeted to women and children and thus tend to be more popular among tax payers who want to ensure that their tax dollars are not wasted, but rather used to "feed children" and reduce social ills such as school dropouts and crime (for example, see de Janvry et al., 1991; Epple and Romano, 2008).

A second potential reason to favor in-kind transfers over cash transfers is that they might have lower de-incentive effects on work and may in fact spur work. A common worry with cash transfers is that they could provide a disincentive to work, particularly if households worry about losing their benefits as their income rises above the eligibility line. It has been argued that in-kind transfer generates fewer labor market distortions and may in fact be labor market enhancing if the provided good is a complement to work. For example, in areas where productivity is low due to nutritional constraints, a food transfer program could ease this constraint. However, the existing evidence thus far does not imply that cash transfers greatly reduce labor market participation (for example, see Alzúa et al., 2013; Banerjee et al., 2015a,b,c), perhaps due to the long duration of benefits and uncertain processes for recertification observed in many developing countries. Furthermore, the cash transfers may also help ease credit constraints for those engaged in the agricultural sector, increasing the productivity of agricultural labor (Gertler et al., 2012).

A third reason in support of in-kind programs is their self-targeting properties (Besley and Coate, 1991; Christian, 2014): by providing a good that the poor differentially value relative to the rich, the poor will apply and the rich will opt out. Again, if money is fully fungible, richer households could simply opt in and sell these goods for the cash, etc. However, we might expect that transition costs and other constraints would imply that in-kind goods are viewed differently than just pure cash. Note that despite this rationale, many in-kind transfer programs are independently targeted prior to distribution, shutting off a channel through which an in-kind program may generate these types of effects. And, while the nonexperimental evidence suggests that in-kind programs are better at selecting the poor (see, for example, Jacoby (1997)), there is little experimental work that compares the magnitude of its targeting properties against different forms of means-tested cash programs.

A final argument in favor of in-kind transfers comes from the idea of missing markets, i.e., if for some reason the market does not provide the good on its own, just providing cash may not be enough. This may be a particular issue in remote areas where high transport costs discourage the spread of products or in disaster or war zones, where food and other products may be in short supply.

4.1.2 Evaluating in-kind programs

One can evaluate an in-kind transfer program in a manner similar to evaluating a cash program, i.e., randomize some areas to receive in-kind transfers (treatment group) and others to not (control group). Then one can collect data to understand if poor households were indeed properly targeted and whether or not they were able to access their entitlement or subsidized products.

However, unlike pure redistributive programs, one may also want to understand whether the theoretical behavioral constraint actually exists, as well as whether the behavioral change that one aimed to induce has occurred. As such, there are two other design features to consider. First, while an evaluation of a redistributive program would be focused on collecting variables to determine whether or not the redistribution has successfully occurred, an evaluation of an in-kind program would additionally require collecting variables to test for the hypothesized behavioral changes.

For example, in evaluating a food transfer program, if the goal is to increase food consumption for children, one would care about collecting detailed modules on food and calories consumed, as well as health indicators for children. However, note two important caveats of this data collection process. First, as Schady et al. (2008) point out, people may hesitate to provide accurate information on surveys if they believe that the surveys are connected to reverification for the program. In which case, one may want to collect variables that are easier to verify: Assets that one can observe, vaccination records on a card, body mass index, other anthropometric variables to assess the health status of children, etc. Second, as the transfer program may be designed to address a number of behavioral changes, one's worry is that in collecting many different variables, we would find some significant impacts just by chance. Thus, one might want to prespecify some of the key hypotheses and outcome variables in advance (Miguel et al., 2014; Olken, 2015).

Second, and perhaps more important, one may also want to consider evaluation strategies that isolate the behavioral constraint. The basic RCT described earlier comparing the effect of a specific program against control areas that do not receive any assistance is important for understanding total program effects, but it does not tell us how the particular constraint (e.g., providing rice versus cash) affects the observed outcomes. Understanding the relevance of specific behavioral constraints can be important, if they imply specific changes to program design.

Therefore, comparing in-kind transfer programs to cash transfer programs can provide useful insights into whether or not constraining behavior is welfare improving. For example, Aker (2014) explores the effect of cash versus food vouchers for displaced households living in an informal camp in the Congo and shows that the level of food consumption was the same regardless of the mechanism since voucher households simply bought food that is relatively easy to sell (e.g., salt). Similarly, the Mexican government took advantage of multiple treatment arms to measure both the effect of in-kind programs to cash and to doing nothing at all. Specifically, they compared three treatments: (1) an in-kind transfer program that gave households 10 different items of food, (2) a cash transfer intended to be of equivalent value, and (3) a control group that did not receive transfers.[18] While the in-kind program distorted consumption of some individual types of food (Cunha, 2014), overall food consumption was similar across both programs (Skoufias et al., 2013) and there was no difference in observed weight among women (Leroy et al., 2013).

Other studies have also collected valuable data that highlight the trade-offs between programs that may satisfy a society's specific goal that an in-kind program may be designed to address (e.g., greater food consumption for kids) versus other important social goals. Hidrobo et al. (2014) compare cash, food, food vouchers, and a control group in Ecuador and find that all three types of programs improve per capita food consumption and caloric intake. However, while food and food vouchers increase calories and diet diversity a bit more relative to cash, the cash program is much easier and cheaper to implement, which may be of real concern for countries with weaker institutional quality.[19]

Similarly, Hoddinott et al. (2014) compare cash versus food transfers in Niger and find that the food transfer program had a larger effect on food consumption and diet variety than cash. However, households were not wasting the funds on temptation goods such as alcohol; they used cash to invest in greater agricultural inputs. Thus, using their estimates,

[18] Note that a particular challenge is ensuring that the in-kind transfer is equivalent to the cash transfer. For example, in the PAL program, the value of the food transfer was 30% more than the cash treatment, since the food basket was based on wholesale prices to the government rather than the prices that consumer pay (see Cunha, 2014 for a discussion of how to make them ex post equivalent).

[19] In an interesting follow-up paper to this experiment, Hidrobo et al. (2013) show that transfers (irrespective of whether food, cash, or vouchers) reduce intimate partner violence.

one can think about how to model the trade-off between the additional utility households receive from spending as they choose, relative to society's utility from the shift in food consumption.

In short, the theory suggested that food, cash, and vouchers could have different effects on household outcomes, but the existing evidence mostly shows that this did not materialize in practice—likely because the particular items of food distributed in these contexts were inframarginal.

4.2 Adding conditions to incentivize behavior

Since the 1990s, it has been increasingly common for many developing country transfer programs to layer a level of "conditionality" onto redistributive transfer programs, e.g., requiring children to go to school and get health checkups for the family to receive its full transfer. The conditions usually aim to correct an underinvestment in the family's well-being that stems from some sort of market failure within the household—parents not internalizing their children's full returns to school, the family not internalizing the benefits of vaccination for society, etc.

CCT programs worldwide have been studied by RCTs—from Mexico's Progresa/Oportunidades to the Philippines' Pantawid Pamilyang Pilipino Program (PPPP)—showing important effects. As with the other programs we have discussed, one can evaluate CCTs by simply randomizing areas to receive or not receive the CCT and then studying the outcomes; this tells us the reduced form effect of having the program relative to no program. In doing so, many of the issues that we have raised earlier—what data to collect, when and how to target, etc.,—would thus be important to think about in this context as well.

However, given the unique nature of the CCT in attempting to correct an investment failure within the family—in addition to redistributing to poor households—more sophisticated analysis can be done to better understand the role of the conditions on household behavior. Later, we discuss three key questions that experimentation can help shed light on (1) What is the impact of the conditions relative to basic redistributive programs?; (2) What should the conditions be?; and finally, (3) How do we enforce them?

4.2.1 Evaluating conditions relative to basic redistributive programs

Evaluations that test a conditional transfer versus no program have difficulty in teasing apart the conditionality mechanism from the liquidity or income effect of the transfer itself. Whether to include conditions in a cash transfer program is an essential design decision. Conditions require a budget to fund them and staff to enforce them. They may also generate selection effects on who chooses to participate.

To isolate the effects of the conditionality, a simple experimental design would have one treatment with a CCT, a second treatment with UCTs, and a control. Thus the second treatment group generates a pure income effect, which allows one to learn whether the conditionality changes behavior, or whether the CCT changes behavior simply through its income effect.

This is the approach taken by Baird et al. (2011) in the Zomba Cash Transfer Program in Malawi. The CCTs performed better at inducing the desired behavior: School dropout rates were lower in the CCT arm than the UCT arm and persisted beyond the program's end. Moreover, attendance rates improved and cognitive ability, mathematics, and English reading comprehension test scores increased for the CCT arm but not the UCT arm. Thus, the conditions had effects on households above and beyond just providing cash. However, the UCT arm had significantly lower marriage and pregnancy rates. This initially may feel like a counterintuitive result since schooling is thought to postpone marriage and lower pregnancy rates. However, the UCT is provided to households even if the teenage girls do not attend school, whereas only those who attend school receive the CCT. Thus, the effect on marriage will depend largely on the relative size of two groups, the group of households that does not attend school under either the UCT or CCT, compared to the group of households that attends school under the CCT but not the UCT. If the income effect on marriage and pregnancy in the first group is large enough (Duflo et al., 2010; Ferre, 2009; Osili and Long, 2008; Ozier, 2011), the UCT will lead to, on net, lower marriage and pregnancy rates. The authors of the Malawi study found this to be true. This dynamic illustrates that the conditions themselves can reduce redistributive aid to the poor—and any associated benefits of the aid—by changing who participates in the program. And, it reinforces the need for careful data collection on ancillary outcomes—e.g., in this case, marriage and pregnancy—to understand the full set of mechanisms through which a program works so that policy-makers can better understand the trade-offs that they would make by implementing one program over another.

Further insights on the trade-off between a UCT and CCT can come through examining what beneficiaries would choose, given a choice. Most programs naturally do not provide such a choice, but providing a choice in an experimental context can help offer insights into which programs would yield higher utility to citizens. For example, individuals could use the conditions in the CCT to generate a personal or family commitment device to engage in future behavior. If respondents were to opt in to such a CCT over a UCT, this is strong evidence of either individual demand for a commitment device (Ashraf et al., 2006; Bryan et al., 2010) or of demand due to family conflict over education or health decisions. Similarly, in Brazil, researchers examined households that were given a choice between a UCT and CCT on school attendance and also tested a subtreatment within the UCT in which households were informed of whether the children were attending school or not (Bursztyn and Coffman, 2012). The parents exhibited a strong preference for the CCT, unless the UCT included monitoring of their children's school attendance, in which case they were content with the UCT. These preferences lend an important insight into the underlying mechanisms of the CCT and suggest that the conditionality in this context was simply a tool for parents to better monitor children. Thus, interventions that improve monitoring and communications between schools and parents may be a better solution than the complicated nature of CCTs over UCTs.

4.2.2 Which conditions should be imposed?

Policy-makers have to make choices about what conditions to impose: Adding conditions adds additional costs of monitoring and as we discussed earlier can have important implications on who participates in programs, potentially screening out the very people whom you want to reach. Thus, experimentation can also help along this dimension, helping to determine which conditions are more impactful and worth focusing on.

In doing so, it is more complicated than just choosing to condition on a single factor, say "schooling." The structure is also important: Conditions can be imposed on inputs or activities (e.g., attendance of children) or outcomes (e.g., test scores) or both. And then, the payment structure must also be defined. For example, Barrera-Osorio et al. (2011) show that whether you simply condition a monthly payment by attendance, or hold back part of the payment and provide it only if the child reenrolls in school, can affect schooling outcomes.

Another type of structure provides conditions for participation but does not financially penalize households if they do not meet them. Thus, this structure essentially provides households with a "nudge." For example, comparing a CCT with a cash transfer program that was "labeled" for schooling (LCT), Benhassine et al. (2015) show that both types of programs improved school participation relative to the control group, but they were not statistically different from each other. The conditionality costs more to implement and so was more expensive relative to the LCT.

As we discussed earlier, applying conditions may lead to a trade-off between the goals of the conditions and the initial goal of redistribution. One legitimate worry is that certain types of conditions may discourage poor households from applying since they find some of conditions too onerous to comply with, thus diminishing the ultimate redistributive goal of these programs. For example, in areas with strong cultural beliefs against vaccines, would requiring vaccines for children reduce the probability that the poorest households participate in the program? One can imagine extensions of the CCT literature with randomization not only for whether the program has conditions but also for the types of conditions in different areas—measuring not only the effect on recipients but also the effect on who applies for the programs and how well the implementers can realistically enforce them.

4.2.3 Enforcing the conditions

The enforcement of the conditionality is critical to examine. Without the enforcement, the program is perhaps theoretically equivalent to an UCT. Politics at the policy level and corruption at the implementation level, both can lead to de facto removal of the conditionality. This occurred, for example, in Ecuador with the *Bono de Desarrollo Humano (BDH)* program. An evaluation of the program thus analyzes the results as if the program is an UCT (Fernald and Hidrobo, 2011; Paxson and Schady, 2010). However, a CCT program that fails to enforce may ultimately generate a different behavioral response than just a pure UCT. For example, in the Ecuador case, many households believed that the funds were indeed conditional, even though in reality they were not.

Enforcement of conditions is an area worthy of further research, but it is not obvious that this is an appropriate randomized trial territory. While one could randomize the enforcement of the program, in an approach similar to what Olken (2007) uses for road building, the political environment that led a government to fail to enforce the conditions may matter, and it may be that merely randomizing enforcement in one setting does provide insights to settings in which enforcement is not viable for political or social reasons. The reason for lack of enforcement is important and cannot be simulated merely through randomization. For example, lack of viability of enforcement could be driven by constituent expectations, local social norms (which both drive the level of enforcement and the treatment effect of the program), or simultaneous policies that interact with the treatment effects of the transfer program.

However, an unenforced condition may be similar to a suggestion or a "nudge." For example, an unenforced condition of school attendance may serve a similar role as the government merely labeling a transfer as an "education support program." Benhassine et al. (2015) examine this question by designing a two-pronged experiment: CCT versus LCT (those two prongs were crossed with a household structure test, providing the program to mothers versus fathers). To understand the underlying mechanisms, the study collected data on process changes and also used a multiarmed experimental design. For example, to understand if the conditionality (or labeling) leads to changes by signaling information about returns to education, researchers collected data on parental beliefs about returns to education (no change was found). Attendance (conditional on enrollment) increased after CCTs and LCTs, leading to an increase in time spent on studying, at school, and traveling to school and a decrease in leisure and productive labor (but not a decrease in chores); this suggests the barrier to schooling was due more to a lack of student interest (thus drawing a similar conclusion on mechanisms as the Brazilian study mentioned earlier (Bursztyn and Coffman, 2012). The multiarm experimental design then tested explicitly the marginal benefit of the condition, over and beyond a merely labeled transfer, and the study found no additive effect beyond the label.

5. MARKET FAILURES PREVENTING ASSET ACCUMULATION

Most of the issues we discussed thus far focus on providing short-term relief through redistributive aid or insurance to households, with some programs paying attention to ensuring that the money is spent on goods and services that benefit households in ways beyond an increase in cash or liquidity. However, certain forms of antipoverty programs also aim to address underlying mechanisms that may be creating poverty traps to improve long-term income for the poor or to provide mechanisms for individuals to build long-term financial assets for when they are elderly. These are, indeed, more complicated challenges: If the underlying market frictions are beyond credit and savings market constraints, then the solution will require more than redistribution. If redistribution alone is employed, it may provide important short-run benefits but may ultimately act as more of a band aid on immediate symptoms without helping individuals achieve a sustained increase in income.

5.1 Building productive assets

For the past 30 years, microcredit has been a leading development policy in the fight to reduce poverty. Unfortunately, seven recent randomized trials have shown that microcredit,[20] while it does provide important benefits, does not improve long-term income, on average, for participants in its current form (for a review, see Banerjee, Karlan, and Zinman, 2015; the seven randomized trials are Angelucci et al., 2015; Attanasio et al., 2015; Augsburg et al., 2015; Banerjee et al., 2015a,b,c; Crépon et al., 2015; Karlan and Zinman, 2011; Tarozzi et al., 2015). In addition, Meager (2015) aggregates the microdata across these studies and uses Bayesian hierarchical models to show that the effect of microcredit on household profits is likely very small and that the effects from each individual study site are reasonably informative for each other. This suggests that either credit constraints are not driving stagnated growth for the poor, the current designs of microcredit programs do not fully address credit constraints, or microcredit may not work without changes in other conditions that also generate market failures for the poor. Multisite studies provide tremendous opportunities for such analysis, e.g., through building more robust theories and then examining, using appropriate statistical tools, how well results from multiple sites fit broad theoretical frameworks.

Indeed, the poor do face multiple failures at the same time that may hinder long-run investment and income growth. But we typically observe programs tackling one problem at a time, rather than a coordinated system. Poverty alleviation policy, as with many government programs, often operates in silos. One silo, described previously and typically managed under the umbrella of "social protection," focuses on redistribution policies through either CCTs or UCTs. A second silo, often managed under the ministries of trade or agriculture, focuses on livelihood support, such as the transfer or a productive asset or agricultural input, alongside some training. A third silo, financial inclusion, is

[20] Note that this lesson was only learned by running similar experiments across different project sites and countries, showing how valuable replication studies can be in changing perspective.

often administered through for-profit or nonprofit (and sometimes subsidized) financial institutions. But, naturally, the causes of poverty may be multifaceted. Thus, uncoordinated programs across different ministries may fail to provide the right bundle of interventions that a household would need to improve their living standard. This lack of coordination in itself poses an interesting research agenda to understand the impacts of these multiapproach programs.

One recent example of such a program is the "graduation" approach—an integrated, multi-faceted program with livelihood promotion at its core that aims to "graduate" individuals out of extreme poverty and onto a long term, sustainable higher consumption path. BRAC, the world's largest nongovernmental organization, has scaled-up this program in Bangladesh (Bandiera et al., 2016), while NGOs around the world have engaged in similar livelihood-based efforts. Six randomized trials across the world (Ethiopia, Ghana, Honduras, India, Pakistan, and Peru) found that the integrated multifaceted program was "sufficient" to increase long-term income, where long term is defined as 3 years after the productive asset transfer (Banerjee et al., 2015a,b,c). The results from the pooled analysis across all six countries found that the program led to sustainable and significant impacts in 10 out of 10 categories of impact. Using an index approach to account for multiple hypotheses testing, positive impacts were found for consumption, income and revenue, asset wealth, food security, financial inclusion, physical health, mental health, labor supply, political involvement, and women's decision-making after 2 years. After a third year, the results remained the same in 8 out of 10 outcome categories (with point estimates falling to below statistical significance for physical health and women's empowerment). Furthermore, the West Bengal site found even larger treatment effects after 7 years (Banerjee et al., 2016). The pattern of results are strikingly similar to the Bangladesh study (Bandiera et al., 2016).

These results are promising in that they show that a sufficient set of interventions is capable of alleviating poverty sustainably and are thus important for policy. They should whet the appetite, both for a more theoretically grounded understanding of exactly which market failures led to a poverty trap, as well as a more practically grounded understanding of whether all of the interventions were truly necessary or if certain components could be removed. In the event that some components are unnecessary, costs could be lowered considerably, allowing the program to reach more people using the same budget. Returning to the theme of this paper, there are two complementary methods to tackle testing the important mechanisms behind the theory, and success or failure, of these programs: data and experimental design.

The ideal method, if unconstrained by budget and organizational constraints, is a complex experimental design that randomizes all permutations of each component. The productive asset transfer, if the only issue were a credit market failure, may have been sufficient to generate these results, and if no other component enabled an individual to accumulate sufficient capital to acquire the asset, the transfer alone may have been a necessary component. The savings component on the other hand may have been a substitute for the productive asset transfer, by lowering transaction costs to save and serving as a behavioral intervention, which facilitated staying on task to accumulate savings. Clearly, it is not realistic in one setting to test the necessity or sufficiency of each component and interaction across components: Even if treated simplistically with each component either present or not, this would imply $2 \times 2 \times 2 \times 2 = 16$ experimental groups.

Data can also provide important insights, even absent experimental design variation. Take the savings component, for example. For the savings component to be either a necessary or sufficient component, presumably an increase in the flow of savings must be observed (but not necessarily the stock, since withdrawals for investment purposes may bring the stock back down). The evidence from the graduation programs shows widely varying impacts on savings, far more than the results of the program itself. For example, in the most extreme case, savings increased in Ethiopia by purchasing power parity (PPP) US$707[21] compared to only PPP US$17 in Ghana. This suggests that savings may be an important component but is neither a necessary nor sufficient component for some level of success.

Several studies have tackled pieces of the puzzle. The way forward is going to be the development of a mosaic of these studies that tests each component, but also includes sufficient contextual and market variations that it can help set policy for a myriad of countries and populations. For example, in a postconflict setting in Uganda, an NGO-led program provided youth groups with training and cash (US$150) toward nonagricultural self-employment activity and found a 57% increase in business assets, 17% increase in work hours, and 38% increase in earnings 4 years after the cash grants (Blattman et al., 2014a). This program differs from the previously mentioned graduation programs in three potentially important and illuminating dimensions: postconflict versus nonpostconflict, youth versus general population of the extreme poor, group-level intervention versus household level, and no inclusion of ancillary components such as life coaching, savings, and health care. The first two differences speak to the applicability of the program to alternative sample frames and settings, whereas the third and fourth program variations suggest that either the driver of the impact of the program lies with the cash grants and training not the other components or that the group-level aspect improves the impact and effectively substitutes for the other components.

A second study in Uganda sheds insight into the value of the group-level intervention, as it randomly varies the group aspect of the intervention, as well as the intensity of supervision (Blattman et al., 2014b). These programs, as with the earlier Uganda program, differ from the graduation studies in that they do not include savings, health and

[21] Although note that the Ethiopia program design included a much stronger push for savings than the other programs, with savings put forward as almost a "mandatory" component, even though there was no consequence if households did not save.

life-coaching components, and are focused on enterprise development (rather than animal husbandry, the dominant livelihood in the graduation studies).

Thus, the initial studies discussed have established a base case that there exists a sufficient intervention package that increases long-term income. We highlight four lines of inquiry to understand more about the underlying mechanisms. First, long-term impacts are critical for assessing whether the short-run interventions actually addressed the underlying problems or rather just lasted a bit longer than a cash transfer. For example, graduation programs typically last 2 years while the graduation studies cited earlier measured impacts 3 years after the assets were transferred. If the household visits were a critical component in driving the observed impacts, longer-term measurement would be important to assess whether the behavioral changes motivated by the household visits persisted for more than just 1 year after the household visits ceased.

Second, as some of the previous studies have begun to do, more work is needed to tease apart the different components: asset transfer (addresses capital market failures), savings account (lowers savings transaction fee), information (addresses information failures), life-coaching (addresses behavioral constraints, and perhaps changes expectations and beliefs about possible return on investment), health services and information (addresses health market failures), consumption support (addresses nutrition-based poverty traps), etc. There will be no simple answer to the aforementioned queries, but further work can help isolate the conditions under which each of these components should be deemed necessary to address. And furthermore, for several of these questions, there are key open issues for *how* to address them; for example, life coaching can take on an infinite number of manifestations. Some organizations conduct life coaching through religion, others through interactive problem solving, and others through psychotherapy approaches (Bolton et al., 2003, 2007; Patel et al., 2010). Much remains to be learned not just about the promise of such life-coaching components, but how to make them work (if they work at all).

Third, general equilibrium effects should be considered, particularly as the programs are taken to scale. Here, the first task is to be more specific in data collection, as general equilibrium effects encompass a wide variety of indirect effects, such as price of transferred assets; spillovers from explicit sharing of granted resources; and increased economic activity from increasing the poor's wealth. A typical experimental design would either randomize across and within villages (assuming that the village is the boundary for generating general equilibrium effects) or for some issues, examining spillovers to nonparticipants in treatment versus control (as in Angelucci and De Giorgi, 2009).

Fourth, important lessons can be learned from understanding the consumption path taken by households after participating in these programs. The graduation program, for example, found important and cost-effective, but still modest, increases in long-term consumption. This finding suggests that households are not caught in an extreme poverty trap, where one simply needs to get households over a particular hump and they will immediately converge to the equivalent of the middle class. Further work is needed to understand the long-term dynamics of such programs and what can be done to further increase income mobility.

5.2 Building long-term financial assets—pensions

Noncontributory pension programs are an important form of social protection in many developing countries—such as Brazil, South Africa, India, etc.,— and given the shift in demographics and the risk of poverty for the elderly (UNDESA, 2013), they are likely to grow in importance. The programs vary in shape and form: Rofman et al. (2015) compare pension programs across 14 different Latin American countries, showing differences in payment sizes, in timing of payment, in whether the pensions are targeted, etc. While to the best of our knowledge, there are few experimental studies of pension programs in developing countries.[22] RCTs can help us understand how differences in these design choices affect the labor market choices of working-age adults, retirement age, saving patterns, and how funds are used within the household.

6. IDEAS ONLY GO SO FAR: IMPLEMENTATION MATTERS TOO

A transfer program may look like a winner on paper but may be a total flop in practice if the implementation is haphazard. Some of this may be purely administrative, e.g., ensuring that the right number of staff is hired, and that they are properly trained and motivated. This may require incentives to not shirk on the job, as well as to not engage in bad behaviors, e.g., siphon off funds or food, or reallocate the funds to friends or political supporters rather than those who are most in need.

Therefore, in designing randomized evaluations of antipoverty programs, it is also important to think about whether, theoretically, a particular aspect of the implementation is likely to be particularly vulnerable to problems. There are two types of variations one can think about: (1) evaluations that vary the underlying structure of the program and (2) evaluations that layer on complementary actions that can be undertaken to improve program implementation given a fixed program design.

Experimentally varying the core elements of the underlying structure of a program is challenging, especially if aspects of the program have been written into law. However, the details of the underlying structure—from who should implement the program to how one should make the transfers—may matter tremendously, affecting the level of leakages and corruption, the targeting, the costs for beneficiaries to access the program, and potentially how beneficiaries spend their entitlements. For example, there is an extensive work, in general, exploring how officials' incentives affect their work output,

[22] Although as of the time of this chapter, there are several exciting ongoing studies in Chile and India.

but to our knowledge less work on how the incentives provided to officials affects transfer program delivery. Similarly, there is a strain of research that shows that the type of person recruited may affect government efficiency (see, for example, recent empirical evidence from Ashraf et al., 2014; Hanna and Wang, 2014), but less specifically on how changing who is selected to implement transfer programs affects the ultimate outcomes of households. For future research, one could vary the salary and incentive structure (amount and conditions) for current workers—and during the recruitment of new workers—to explore how it affects targeting and delivery.

One important question is whether governments should even directly implement these programs, or whether they should contract out delivery mechanisms. Banerjee et al. (2014) experimentally vary whether local officials distribute a government-run subsidized rice program or whether private citizens also bid for the right to run the program. They find that the bidding reduces the price—markup that citizens pay, without reducing quality. Follow-up work can include testing out different ways of contracting out, from changing who is eligible to bid to how the bidding process occurs to how new implementers are reevaluated. Moreover, the bidding process in that paper focuses on local government provision (i.e., at the village level). Future research could also help shed light on whether the procurement process should be done at that local level where individuals possess local information about how to get things done in that village, or should it be done at a higher level of government (e.g., district or province level) where one may also benefit from economies of scale.

A nice series of recent papers tests whether the nature of the delivery mechanism in itself affects outcomes. For example, Aker et al. (2011) experimentally test for the impact of providing cash versus mobile money in a short-run transfer program in Niger. An innovative feature of their experimental design was to also have a treatment group that simply got cash *and a cell phone* and to net out potential effects of having a cell phone more generally from mobile money. Mobile money not only reduced the nonprofit's distribution costs but it also reduced the households' costs to pick up their entitlement. This second feature of mobile money may be particularly important if we believe high transaction costs induce beneficiary households "leave money" on the table (Currie and Gahvari, 2008). Importantly, they also showed that spending patterns changed due to mobile money, hypothesizing that it also conferred greater privacy over one's finances. Further testing these mechanisms in the context of larger government programs to understand longer-run effects would be an important extension of this work: For example, would the ease and potential secrecy of payments attract richer people to apply for these types of transfer programs? In the long run, would local officials who may have previously siphoned off cash during disbursements find other ways to "tax" citizens who now receive cash directly via mobile money?

An ambitious project by Muralidharan et al. (2014) also aims to address some of these types of questions. They evaluated the impact of biometrically authenticated payments infrastructure ("smart cards") on beneficiaries of employment (NREGA) and pension (SSP) programs in Andhra Pradesh, India. The smart cards changed both how households collected their payments, as well as who was in charge of the cash distribution (as banks and technology service providers managed the new cash disbursal system). The state was rolling out the program across its 158 subdistricts, so the authors randomized which subdistricts were converted first. Following the introduction of the program, not only did the time it took beneficiaries to collect a payment fall, but also the delay in receiving the payment was reduced by almost 30%. The ease of payment induced households to actually work more. Households, thus, earned more, while payments to officials remained the same—hence, leakages fell quite dramatically. Similarly, an RCT conducted by Banerjee et al. (2014) shows that by simply asking local officials to input all of the names of the people who participated in NREGA into a database system to receive the funds transfer—i.e., increasing informational requirements for releasing funds and reducing the administrative tiers in the flow of funds process—led to a stark decline in leakages of public transfers, with no corresponding decrease in actual NREGA work.

Experimentally testing complementary programs that are layered on top of existing programs can also be important in improving the delivery of social protection programs. These programs do not necessarily require changing the existing program rules or functioning but instead provide additional information or services to help citizens better access their entitlements. For example, one could test how increased information on eligibility and program rules affects overall program leakages. Ravallion et al. (2013) do this: They experimentally vary whether beneficiaries see a half hour video on their entitlements under NREGA. They show that this form of information has very little impact on employment. Given that the form and level of information may matter, one may also test between varying types of information: for example, Banerjee et al. (2014) show that a card that informs households of their eligible status and entitlements reduce leakages in a subsidized rice program, and that making the card information public within the village has even larger impacts. Moreover, one can imagine experiments designed to test how providing households with direct help with their paperwork when applying affects who enrolls.

Questions about implementation at scale also relate to critical questions about the role of randomized trials given how they are often conducted in developing countries. For example, there is a broader debate about whether we would observe similar program results in NGO and government settings, given differences in implementation capacity between two (see, for example, Bold et al., 2013; Dhaliwal and Hanna, 2014). Of course, if one is evaluating a program with an NGO that will be scaled up by that NGO or similar ones, we may not particularly care if the program would look different if run by the government. However, often times we may also want to understand how evaluations with NGOs would differ in government and vice versa. Naturally, the fundamental problem here is of generalizability and sample size: a comparison of any one NGO to any one government only compares that of NGO to that government. NGOs are not a monolithic set of institutions, and neither are governments. Thus it may be wrong to ask whether a government is better, or worse, at implementing than an NGO, and be more appropriate to ask whether "an" institution with certain specific characteristics or in certain specific cultural or political environments will be better at implementing than an institution with a different set of specific characteristics or environmental factors.

Treatment effects may depend on institutional type (government or NGO) for two broad reasons: behavioral responses and implementation efficiency. In terms of behavioral responses, treatment effects may depend upon the legitimacy of who delivers the program. In the specific case of social protection programs, how we expect households to respond to a particular transfer of food or cash is unlikely to change based on who is distributing it. But, how people respond to the specifics of the program may matter. For example, in the study in Morocco with labeled cash transfers (Benhassine et al., 2015), it could be that such labels only work from trusted and well-known institutions. Again, this is less of an issue over whether the NGO or government is delivering the service but about the overall level of legitimacy of the institution. Thus, one interesting design would be to see if the response to the nudges changes when households are randomized to receive more or less information on how legitimate the organization has been in implementing these programs in the past.

In terms of implementation efficiency, one can also imagine that different types of organizations may have different strengths and weaknesses in terms of the types of programs that they can deliver. Suppose, for example, that citizens would be indifferent between cash and an in-kind transfer if both were implemented perfectly. However, one type of organization has strength in reducing the leakages in the delivery of cash relative to a second organization, and the second organization is better at reducing leakages in the in-kind transfer. Thus, citizens would ultimately prefer different types of transfers from different types of organizations due to the relative differences in leakages. Again, this preference may be symptomatic of a difference between government and an NGO but speaks to larger differences in the relative implementation abilities of different organizations and how can one improve upon their weaknesses.

7. CONCLUSION: KEY AREAS FOR FURTHER WORK

Since the innovative and instrumental randomized evaluation of Progresa in Mexico, there has been a burst of important and exciting experiments in this area. This has greatly informed our understanding of what can "work" in trying to redistribute to the poor, as well as what can reduce both behavioral constraints and market failures.

So, then the question becomes, where should we focus our research efforts next? We highlight three areas for further work, aside from those discussed earlier: interactions of demand and supply, long-term effects, and general equilibrium effects.

7.1 Key areas for further work

7.1.1 Interactions of demand and supply

A vital question is how transfer programs work across different contexts. For example, Galiani and McEwan (2013) document that the effect of the Honduran PRAF CCT program was much larger in the two poorest strata, with the effect not being statistically significant in the three richer areas.

This question is similar to the one at the center of the debate over who implements (e.g., a nonprofit or government): If one wants to understand how a program will work in a specific context, we may not care whether the evaluation findings are portable to another context. But if we want to understand whether a program would have similar results in another area, or if the results will change with policy changes in the current area, it is important to understand how the theoretically important underlying features of an area impact outcomes.

More broadly, there are lots of unanswered questions about the interaction of transfer programs with existing conditions, particularly supply-side conditions: For example, How does school quality or health-care availability affect the adherence to CCT conditions? Do food or other in-kind transfers work better than cash in areas with more limited food supplies? Do transfers facilitate access to finance by reducing risk to lenders? And so forth.

To answer these kinds of questions, one would ideally not only vary the introduction of a transfer program but would also cross this with an experimental change in a supply-side feature. For example, to isolate how increased health-care availability affects the adherence to CCT health conditions, one would randomize areas to four treatments: a pure control, CCT only, an increase in nurses only, and CCT and an increase in nurses.

An extension of this would be to test the effectiveness of different types of transfer programs under different conditions: For example, we might think that a UCT may be more effective at redistributing to the poor than a CCT in areas where there is limited health availability, since the inability to adhere to the conditions may scare off or reduce payments to the poor. Thus, rather than having just a pure control, one may want to compare CCTs to UCTs across areas with and without the induced increases in nurse availability.

7.1.2 Long-term effects

There is a tension between trying to measure a program's long-run impacts versus scaling up a "working" program to the control group. However, long-run impacts are important to measure, especially if there are reasons to believe that a program's impacts may evolve differently as time goes on (and potentially have general equilibrium effects as we discuss later).

For example, while we know quite a bit about the short-run impacts of different targeting methods, we know less about their relative long-run effects. Using quasiexperimental variation, Camacho and Conover (2011) show that Colombia's targeting system was manipulated over time, as local officials better learned the rules of the game. One can imagine experimentally varying different targeting methods across different locations, and then repeating the same method in each respective location during the recertification process, to determine whether the relative efficacy of different methods change as both households and officials learn the systems over time.

Similarly, there are many questions about the long-run impacts of the transfers themselves: What happens to households after the transfers are complete? For example, Did CCTs achieve the goal of changing the outcomes the next generation, i.e., Did the children who attended school for longer, or had improved test scores, ultimately do better in the labor market? In the case of the graduation studies discussed earlier, Bangladesh and West Bengal, India sites have followed households for 7 years, and found that the positive treatment effects increased from 3–7 years (Banerjee et al., 2016).

7.1.3 General equilibrium effects

With relatively large sums being distributed, antipoverty programs may have broader effects than one initially expects, with these effects being potentially quite large. These effects can take various forms. Antipoverty programs can affect insurance and lending markets within villages (Angelucci and de Giorgi, 2009), natural resource demand (Gertler et al., 2013; Hanna and Oliva, 2015), and labor markets (Muralidharan et al., 2016). Some of the effects could be positive, while others can be negative. While we touched on the idea of general equilibrium effects earlier, it is important enough of a topic to warrant its own section here.

Importantly, the general equilibrium effects may differ by type of transfers. For example, Cunha et al. (2013) document different effects on prices of consumer goods when villages have been randomized to cash versus in-kind transfer programs, particularly in remote areas. The general equilibrium effects may also vary across contexts: For example, CCTs induce the positive peer effects on the schooling outcomes of ineligible children in Mexico's Progresa (Bobonis and Finan, 2009; Lalive and Cattaneo, 2009) but no effects on ineligible children in the Honduran PRAF (Galiani and McEwan, 2013). At the more extreme, Barrera-Osorio et al. (2011) find negative spillovers: Siblings (particularly sisters) of CCT recipients are less likely to attend school and more likely to drop out.

While there have been a few studies, including those discussed earlier, that have tried to capture broader effects, this is still an area where our understanding is relatively sparse and where there is a need for further research. However, given the multitude of types of general equilibrium effects that may be possible, this is a case where we are particularly worried about multiple hypothesis testing. Careful theory-detailing predictions, coupled with prespecified hypotheses, may be important in building a robust model that successfully predicts outcomes in new settings, thus properly guides policy-making.

Identifying spillover effects should be built into the research plan when plausible and viable. Three basic approaches have been employed: (1) through experimental design: randomizing the density of treatment within a geographic area (or within any unit within which one expects there to be spillovers or general equilibrium effects), (2) through data collection on ineligibles: this is strengthened when combined with the first, but even on its own can shed important insights (Angelucci and de Giorgi, 2009), and (3) through data collection on process changes: for example, collecting specific data on informal transfers, credit and savings could identify behaviors that indicate the presence of general equilibrium effects.

7.2 Final thoughts

Putting these elements together poses its own challenges. Naturally, there is no simple diagnostic that assesses which markets are missing for a society, to then provide an easy prescription for which of the earlier programs to implement. And, on the other end of the spectrum, it is simply not practical to implement at scale a program that assesses for each individual what constraints they face and then provides the exact program that targets their particular situation. The policy challenge lies in how to find the policy that balances the operational constraints of scale with the targeting constraints of both identifying the poorest and minimizing the false positives, i.e., policies targeting an issue that are not relevant for a household or community. National social protection strategies should think holistically: How do specific policies interact with each other as either complements or substitutes? What populations are critically missing from existing policies? and how well do existing policies improve long-term outcomes so as to reduce the eventual tax burden on society?

REFERENCES

Aker, J.C., 2014. Comparing Cash and Voucher Transfers in a Humanitarian Context: Evidence from the Democratic Republic of Congo.

Aker, J.C., Boumnijel, R., McClelland, A., Tierney, N., 2011. Zap it to Me: The Short-Term Impacts of a Mobile Cash Transfer Program. Center for Global Development. Working Paper No. 268.

Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Purnamasari, R., Wai-Poi, M., 2013. Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia. Working Paper 18798. National Bureau of Economic Research. http://www.nber.org/papers/w18798.

Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Tobias, J., 2012. Targeting the poor: evidence from a field experiment in Indonesia. Am. Econ. Rev. 102 (4), 1206–1240. http://dx.doi.org/10.1257/aer.102.4.1206.

Alatas, V., Banerjee, A., Hanna, R., Olken, B., Purnamasari, R., 2016. Self-targeting: evidence from a field experiment in Indonesia. J. Polit. Econ. 124 (2), 371–427.

Alzúa, M.L., Cruces, G., Ripani, L., 2013. Welfare programs and labor supply in developing countries: experimental evidence from Latin America. J. Popul. Econ. 26 (4), 1255–1284.

Angelucci, M., De Giorgi, G., 2009. Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption? Am. Econ. Rev. 99 (1), 486–508. http://dx.doi.org/10.1257/aer.99.1.486.

Angelucci, M., Karlan, D., Zinman, J., 2015. Microcredit impacts: evidence from a randomized microcredit program placement experiment by compartamos banco. Am. Econ. J. 7 (1), 151–182. http://dx.doi.org/10.1257/app.20130537.

Ashraf, N., Bandiera, O., Lee, S.S., 2014. Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery. Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Ashraf, N., Karlan, D., Yin, W., 2006. Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines. Q. J. Econ. 121 (2), 673–697.

Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E., Harmgart, H., 2015. The impacts of microfinance: evidence from joint-liability lending in Mongolia. Am. Econ. J. 7 (1), 90–122. http://dx.doi.org/10.1257/app.20130489.

Augsburg, B., De Haas, R., Harmgart, H., Meghir, C., 2015. The impacts of microcredit: evidence from Bosnia and Herzegovina. Am. Econ. J. 7 (1), 183–203. http://dx.doi.org/10.1257/app.20130272.

Baird, S., McIntosh, C., Özler, B., 2011. Cash or condition? Evidence from a cash transfer experiment. Q. J. Econ. 126 (4), 1709–1753. http://dx.doi.org/10.1093/qje/qjr032.

Baker, J.L., Grosh, M.E., 1994. Poverty reduction through geographic targeting: how well does it work? World Dev. 22 (7), 983–995. http://dx.doi.org/10.1016/0305-750X(94)90143-0.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., Sulaiman, M., 2016. Labor Markets and Poverty in Village Economies. LSE Working Paper. http://sticerd.lse.ac.uk/dps/eopp/eopp43.pdf.

Banerjee, A., Duflo, E., Chattopadhyay, R., Shapiro, J., 2016. Long Term Impact of a Livelihood Intervention: Evidence from West Bengal (Working Paper).

Banerjee, A., Duflo, E., Glennerster, R., Kinnan, C., 2015a. The miracle of microfinance? Evidence from a randomized evaluation. Am. Econ. J. 7 (1), 22–53. http://dx.doi.org/10.1257/app.20130533.

Banerjee, A., Hanna, R., Kyle, J., Olken, B., Sumarto, S., 2014. Information Is Power: Identification Cards and Food Subsidy Programs in Indonesia (Working Paper).

Banerjee, A., Karlan, D., Zinman, J., 2015b. Six randomized evaluations of microcredit: introduction and further steps. Am. Econ. J. 7 (1), 1–21.

Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. J. Political Econ. 101, 274–298.

Banerjee, A.V., Hanna, R., Kreindler, G., Olken, B.A., 2015c. Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs Worldwide. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2703447.

Bardhan, P.K., Mookherjee, D., 2000. Capture and governance at local and national levels. Am. Econ. Rev. 90 (2), 135–139. http://dx.doi.org/10.1257/aer.90.2.135.

Barrera-Osorio, F., Bertrand, M., Linden, L., Perez-Calle, F., 2011. Improving the design of conditional transfer programs: evidence from a randomized education experiment in Colombia. Am. Econ. J. 3 (2), 167–195.

Barrera-Osorio, F., Filmer, D., 2013. Incentivizing Schooling for Learning: Evidence on the Impact of Alternative Targeting Approaches. World Bank Policy Research Working Paper, no. 6541.

Beath, A., Christia, F., Enikolopov, R., 2013. Do elected councils improve governance? Experimental evidence on local institutions in Afghanistan. MIT Political Sci. Dep. Res. Pap. 2013 (24).

Beaman, L., Karlan, D., Thuysbaert, B., Udry, C., 2014. *Self-Selection into Credit Markets: Evidence from Agriculture in Mali* (Working Paper No. 20387). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w20387.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., Pouliquen, V., 2015. Turning a shove into a nudge? a 'labeled cash transfer' for education. Am. Econ. J. 7 (3), 86–125.

Berry, J., 2014. Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India (Working Paper).

Besley, T., Coate, S., 1991. Public provision of private goods and the redistribution of income. Am. Econ. Rev. 81 (4), 979–984.

Besley, T., Coate, S., 1992. Workfare versus welfare: incentive arguments for work requirements in poverty-alleviation programs. Am. Econ. Rev. 82 (1), 249–261.

Blattman, C., Fiala, N., Martinez, S., 2014a. Generating skilled self-employment in developing countries: experimental evidence from Uganda. Q. J. Econ. 129 (2), 697–752.

Blattman, C., Green, E., Annan, J., Jamison, J., 2014b. The Returns to Cash and Microenterprise Support Among the Ultra-poor: A Field Experiment. Columbia University. Working Paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2439488.

Bobonis, G.J., Finan, F., 2009. Neighborhood peer effects in secondary school enrollment decisions. Rev. Econ. Statistics 91 (4), 695–716.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., Sandefur, J., 2013. Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. Center for Global Development. Working Paper, no. 321.

Bolton, P., Bass, J., Betancourt, T., Speelman, L., Onyango, G., Clougherty, K.F., Neugebauer, R., Murray, L., Verdeli, H., 2007. Interventions for depression symptoms among adolescent survivors of war and displacement in northern Uganda: a randomized controlled trial. JAMA 298 (5), 519. http://dx.doi.org/10.1001/jama.298.5.519.

Bolton, P., Bass, J., Neugebauer, R., Verdeli, H., Clougherty, K.F., Wickramaratne, P., Speelman, L., Ndogoni, L., Weissman, M., 2003. Group interpersonal psychotherapy for depression in rural Uganda: a randomized controlled trial. JAMA 289 (23), 3117–3124.

Bryan, G., Karlan, D., Nelson, S., 2010. Commitment devices. Ann. Rev. Econ. 2 (1), 671–698. http://dx.doi.org/10.1146/annurev.economics.102308.124324.

Bursztyn, L., Coffman, L.C., 2012. The schooling decision: family preferences, intergenerational conflict, and moral hazard in the Brazilian favelas. J. Political Econ. 120 (3), 359–397. http://dx.doi.org/10.1086/666746.

Camacho, A., Conover, E., 2011. Manipulation of social program eligibility. Am. Econ. J. Econ. Policy 3 (2), 41–65.

Cameron, A.C., Miller, D.L., 2010. Robust Inference with Clustered Data. University of California, Department of Economics. Working Papers. http://www.econstor.eu/handle/10419/58373.

Chen, S., Ravallion, M., Wang, Y., 2006. Di Bao: A Guaranteed Minimum Income in China's Cities?, Vol. 3805. World Bank Publications, Washington, D.C.

Chetty, R., Looney, A., 2006. Consumption smoothing and the welfare consequences of social insurance in developing economies. J. Public Econ. 90 (12), 2351–2356. http://dx.doi.org/10.1016/j.jpubeco.2006.07.002.

Christian, P., 2014. The Distributional Consequences of Group Procurement: Evidence from a Randomized Trial of a Food Security Program in Rural India (Working Paper).

Coady, D., 2004. Designing and Evaluating Social Safetey Nets: Theory, Evidence, and Policy Conclusions. Food Consumption and Nutrition. Division Discussion Paper No. 172.

Cohen, J., Dupas, P., 2010. Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. Q. J. Econ. 125 (1), 1–45. http://dx.doi.org/10.1162/qjec.2010.125.1.1.

Covarrubias, K., Davis, B., Winters, P., 2012. From protection to production: productive impacts of the Malawi social cash transfer Scheme. J. Dev. Eff. 4 (1), 50–77.

Crépon, B., Devoto, F., Duflo, E., Pariente, W., 2015. Estimating the impact of microcredit on those who take it up: evidence from a randomized experiment in Morocco. Am. Econ. J. Appl. Econ. 7 (1), 123–150. http://dx.doi.org/10.1257/app.20130535.

Cunha, J.M., De Giorgi, G., Jayachandran, S., 2013. The Price Effects of Cash Versus In-kind Transfers.

Cunha, J.M., 2014. Testing Paternalism: Cash versus In-Kind Transfers. *American Economic Journal: Applied Economics* 6 (2), 195–230. http://dx.doi.org/10.1257/app.6.2.195.

Currie, J., Gahvari, F., 2008. Transfers in cash and in-kind: theory meets the data. J. Econ. Literature 46 (2), 333–383.

de Janvry, A., Fafchamps, M., Sadoulet, E., 1991. Peasant household behaviour with missing markets: some paradoxes explained. Econ. J. 101 (409), 1400. http://dx.doi.org/10.2307/2234892.

de Mel, S., McKenzie, D., Woodruff, C., 2008. Returns to capital in microenterprises: evidence from a field experiment. Q. J. Econ. 123 (4), 1329–1372.

Dhaliwal, I., Hanna, R., 2014. Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India. National Bureau of Economic Research. Working Paper 20482. http://www.nber.org/papers/w20482.
Duflo, E., Dupas, P., Kremer, M., 2010. Education and Fertility: Experimental Evidence from Kenya (Working Paper).
Duflo, E., Gale, W., Liebman, J., Orszag, P., Saez, E., 2006. Saving incentives for low- and middle-income families: evidence from a field experiment with H&R block. Q. J. Econ. 121 (4), 1311–1346.
Dupas, P., Hoffmann, V., Kremer, M., Zwane, A.P., 2013. Micro-ordeals, Targeting, and Habit Formation (Working Paper).
Edmonds, E., Schady, N., 2012. Poverty alleviation and child labor. Am. Econ. J. Econ. Policy 4 (4), 100–124.
Elbers, C., Fujii, T., Lanjouw, P., Özler, B., Yin, W., 2007. Poverty alleviation through geographic targeting: how much does disaggregation help? J. Dev. Econ. 83 (1), 198–213.
Epple, D., Romano, R., 2008. Educational vouchers and cream skimming. Int. Econ. Rev. 49 (4), 1395–1435.
Fernald, L.C.H., Hidrobo, M., 2011. Effect of Ecuador's Cash Transfer Program (Bono de Desarrollo Humano) on Child Development in Infants and Toddlers: a Randomized Effectiveness Trial. Soc. Sci. Med. 72 (9), 1437–1446. http://dx.doi.org/10.1016/j.socscimed.2011.03.005.
Ferre, C., 2009. Age at First Child: Does Education Delay Fertility Timing? the Case of Kenya. Social Science Research Network, Rochester, NY. SSRN Scholarly Paper ID 1344718. http://papers.ssrn.com/abstract=1344718.
Fiszbein, A., Schady, N.R., 2009. Conditional Cash Transfers: Reducing Present and Future Poverty. World Bank, Washington, DC. A World Bank Policy Research Report 47603.
Fudenberg, D., Levine, D., 2006. A dual self model of impulse control. Am. Econ. Rev. 96 (5), 1449–1476.
Galiani, S., McEwan, P.J., 2013. The heterogeneous impact of conditional cash transfers. J. Public Econ. 103 (C), 85–96.
Gentilini, U., Honorati, M., Yemtsov, R., 2014. The State of Social Safety Nets 2014. World Bank Group, Washinton, DC.
Gertler, P.J., Martinez, S.W., Rubio-Codina, M., 2012. Investing cash transfers to raise long-term living standards. Am. Econ. J. Appl. Econ. 4 (1), 164–192.
Gertler, P., Shelef, O., Wolfram, C., Fuchs, A., 2013. How Pro-poor Growth Affects the Demand for Energy. National Bureau of Economic Research. Working Paper 19092. http://www.nber.org/papers/w19092.
Glewwe, P., Park, A., Zhao, M., 2014. A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China. University of Minnesota Center for International Food and Agricultural Policy. Working Paper WP12-2.
Grenier, J., Pattanayak, C.W., 2011. Randomized evaluation in legal assistance: what difference does representation (offer and actual use) make. Yale LJ 121, 2118.
Hanna, R., Oliva, P., 2015. The effect of pollution on labor supply: evidence from a natural experiment in Mexico city. J. Public Econ. 122, 68–79.
Hanna, R., Wang, S.-Y., 2014. Dishonesty and Selection into Public Service: Evidence from India.
Hidrobo, M., Hoddinott, J., Peterman, A., Margolies, A., Moreira, V., March 2014. Cash, food, or vouchers? evidence from a randomized experiment in northern Ecuador. J. Dev. Econ. 107, 144–156. http://dx.doi.org/10.1016/j.jdeveco.2013.11.009.
Hidrobo, M., Peterman, A., Heise, L., 2013. The Effect of Cash, Vouchers and Food Transfers on Intimate Partner Violence: Evidence from a Randomized Experiment in Northern Ecuador. International Food Policy Research Institute, Washington, DC. https://www.wfp.org/sites/default/files/IPV-Hidrobo-Peterman_Heise_IPV%20Ecuador%203%2028%2014.pdf.
Hoddinott, J., Sandström, S., Upton, J., 2014. The Impact of Cash and Food Transfers: Evidence from a Randomized Intervention in Niger. Available at: SSRN 2423772. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2423772.
Jacoby, H.G., 1997. Self-selection and the redistributive impact of in-kind transfers: an econometric analysis. J. Hum. Resour. 233–249.

Jessee, C., Prencipe, L., Sherman, D., Banda, A., Ndiyoi, L., Tembo, N., Daidone, S., et al., 2013. Zambia's Child Grant Program: 24-Month Impact Report. American Institutes for Research.
Karlan, D., Zinman, J., 2011. Microcredit in theory and practice: using randomized credit scoring for impact evaluation. Science 332 (6035), 1278–1284. http://dx.doi.org/10.1126/science.1200138.
Kazianga, H., De Walque, D., Alderman, H., 2012. Educational and child labour impacts of two food-for-education schemes: evidence from a randomised trial in rural Burkina Faso. J. Afr. Econ. 21 (5), 723–760.
Kling, J.R., Liebman, J.B., Katz, L.F., Sanbonmatsu, L., 2004. Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-sufficiency and Health from a Randomized Housing Voucher Experiment. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=588942.
Kremer, M., Miguel, E., Thornton, R., 2009. Incentives to learn. Rev. Econ. Statistics 91 (3), 437–456.
Kuziemko, I., Norton, M.I., Saez, E., Stantcheva, S., 2015. How elastic are preferences for redistribution? Evidence from randomized survey experiments. Am. Econ. Rev. 105 (4), 1478–1508. http://dx.doi.org/10.1257/aer.20130360.
Lalive, R., Cattaneo, M.A., 2009. Social interactions and schooling decisions. Rev. Econ. Statistics 91 (3), 457–477.
Leroy, J.L., Gadsden, P., González de Cossío, T., Gertler, P., 2013. Cash and in-kind transfers lead to excess weight gain in a population of women with a high prevalence of overweight in rural Mexico. J. Nutr. 143 (3), 378–383.
Ma, X., Sylvia, S., Boswell, M., Rozelle, S., 2013. Ordeal Mechanisms and Information in the Promotion of Health Goods in Developing Countries: Evidence from Rural China. Rural Education Action Project Working Paper No. 266.
McKenzie, D., Woodruff, C., 2006. Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises. Econ. Dev. Cult. Change 55 (1), 3–42. http://dx.doi.org/10.1086/505725.
Meager, R., 2015. Understanding the Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of 7 Randomised Experiments. Available at: SSRN 2620834. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2620834.
Merttens, F., Hurrell, A., Marzi, M., Attah, R., Farhat, M., Kardan, A., MacAuslan, I., 2013. Kenya Hunger Safety Net Programme Monitoring and Evaluation Component. Impact Evluation Final Report. Oxford Policy Management.
Meyer, B.D., 1995. Lessons from the US unemployment insurance experiments. J. Econ. Literature 33 (1), 91–131.
Mickelwright, J., Nagy, G., 2010. The effect of monitoring unemployment insurance recipients on unemployment duration: evidence from a field experiment. Labour Econ. 17 (1), 180–187.
Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., et al., 2014. Promoting transparency in social science research. Science 343 (6166), 30–31.
Muralidharan, K., Niehaus, P., Sukhtankar, S., 2014. Building State Capacity: Evidence from Biometric Smartcards in India. NBER. Working Paper No 19999.
Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, 2016. General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India. MIMEO.
Nichols, A.L., Zeckhauser, R.J., 1982. Targeting transfers through restrictions on recipients. Am. Econ. Rev. 372–377.
Norton, M.I., Ariely, D., 2011. Building a better America—one wealth quintile at a time. Perspect. Psychol. Sci. 6 (1), 9–12. http://dx.doi.org/10.1177/1745691610393524.
Olken, B., 2007. Monitoring corruption: evidence from a field experiment in Indonesia. J. Political Econ. 115, 200–249.
Olken, B.A., 2015. Promises and perils of pre-analysis plans. J. Econ. Perspect. 29 (3), 61–80.
Osili, Okonkwo, U., Long, B.T., 2008. Does female schooling reduce fertility? evidence from Nigeria. J. Dev. Econ. 87 (1), 57–75. http://dx.doi.org/10.1016/j.jdeveco.2007.10.003.
Ozier, O., 2011. The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis. Working Paper. http://economics.ozier.com/owen/papers/ozier_JMP_20110117.pdf.

Patel, V., Weiss, H.A., Chowdhary, N., Naik, S., Pednekar, S., Chatterjee, S., De Silva, M.J., Bhat, B., Araya, R., King, M., 2010. Effectiveness of an intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomised controlled trial. Lancet 376 (9758), 2086–2095.

Paxson, C., Schady, N., 2010. Does money matter? the effects of cash transfers on child development in rural Ecuador. Econ. Dev. Cult. Change 59 (1), 197–229.

Ravallion, M., 1991. Reaching the rural poor through public employment: arguments, evidence, and lessons from South Asia. World Bank Res. Observer 6 (2), 153–175.

Ravallion, M., Van de Walle, D.P., Dutta, P., Murgai, R., 2013. Testing Information Constraints on India's Largest Antipoverty Program. World Bank Policy Research. Working Paper, no. 6598. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2323980.

Rawls, J., 1971. A Theory of Justice, Rev ed. Belknap Press of Harvard University Press, Cambridge, Mass.

Rofman, R., Apella, I., Vezza, E., 2015. Beyond Contributory Pensions: Fourteen Experiences with Coverage Expansion in Latin America. World Bank Publications. https://books.google.com/books?hl=en&lr=&id=5GOzBQAAQBAJ&oi=fnd&pg=PP1&dq=Beyond+Contributory+Pensions:+Fourteen+Experiences+with+Coverage+Expansion+in+Latin+America&ots=OCI8ZK8PDk&sig=Bza3L1VoqX0fNtcDLvnIl3KrOUY.

Saavedra, J., Garcia, S., 2012. Impacts of Conditional Cash Transfer Programs on Educational Outcomes in Developing Countries. Product Page. http://www.rand.org/pubs/working_papers/WR921-1.html.

Schady, N., Araujo, M.C., Peña, X., López-Calva, L.F., 2008. Cash transfers, conditions, and school enrollment in Ecuador [with comments]. Econ. J. Lat. Am. Caribb. Econ. Assoc. 43–77.

Schultz, T.P., 2004. School subsidies for the poor: evaluating the Mexican progresa poverty program. J. Dev. Econ. 74 (1), 199–250.

Seabright, P., 1996. Accountability and decentralisation in government: an incomplete contracts model. Eur. Econ. Rev. 40 (1), 61–89.

Singer, P., 1997. The drowning child and the expanding circle. In: New Internationalist, April, 28–30.

Skoufias, E., Unar, M., de Cossio, T.G., 2013. The poverty impacts of cash and in-kind transfers: experimental evidence from rural Mexico. J. Dev. Eff. 5 (4), 401–429.

Tarozzi, A., Desai, J., Johnson, K., 2015. The impacts of microcredit: evidence from Ethiopia. Am. Econ. J. Appl. Econ. 7 (1), 54–89. http://dx.doi.org/10.1257/app.20130475.

The Kenya CT-OVC Evaluation Team, 2012. The impact of Kenya's cash transfer for Orphans and vulnerable children on human capital. J. Dev. Eff. 4 (1), 38–49.

UNDESA, 2013. World Population Ageing Report. UN Department of Economic and Social Affairs.

UNHCR, 2014. Statistical Online Population Database. United Nations High Commissioner for Refugees.

Van den Berg, G.J., Van der Klaauw, B., 2006. Counseling and monitoring of unemployed workers: theory and evidence from a controlled social experiment. Int. Econ. Rev. 47 (3), 895–936.

Vermeersch, C., Kremer, M., 2004. Schools Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation. The World Bank.