Designing Anti-Poverty Programs: Experimental Lessons

Rema Hanna

Harvard University

Dean Karlan

Yale University

DRAFT AND INCOMPLETE

Abstract

Anti-poverty programs that aim to help individuals establish a sustainable income source come in many varieties, and range from multi-faceted complex, even dynamic, interventions to simple cash transfers. The more complex and multi-faceted intervention pose a challenge for research. While straightforward to answer "whether" something works, it is far less straightforward to answer why it works. We will discuss these challenges, and then focus on how careful experimentation, typically employing multiple-treatment arms, complemented with thoughtful data collection and analysis, can help overcome these obstacles. We will discuss issues along several dimensions, for example targeting of the program (and how that affects not just who participates but their level of engagement in the program), who implements the program, and the conditionality of transfers. Ultimately, with a stronger understanding of why programs work, or do not, one can make stronger policy prescriptions.

I. Introduction

Social protection programs are already common in many developing countries and we expect their importance in policy to further rise over time. In low-income countries, more than one billion individuals are enrolled in at least one safety net program. Nonetheless, coverage is incomplete: in the poorest countries, where 47 percent of the population is extremely poor, fewer than 10 percent have access to formal social protection programs (Gentilini et al, 2014). Poor countries spend a smaller percentage of GDP on social protection than do rich ones and, therefore, we expect that poor countries' social protection expenditures will rise as a proportion of GDP as they get richer. Thus, understanding the impacts of different forms of social protection programs in developing countries is of key policy importance.

Most social protection (also referred to as "safety net") programs aim not only to transfer resources to the poor for redistributive purposes, but also to address the potential short- and long-term consequences from income shocks. The poor face many risks, such as unexpected health costs, agricultural damage or job losses. Without fully functioning and formal credit markets, households may be unable to borrow to smooth consumption. Informal credit and insurance markets provide another avenue to do so, but they are often underperforming 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.

Increasingly though, governments and non-profits are adopting more nuanced social protection programs, aimed not only to supplement consumption in hard times, but to address the underlying market failures that lead income shocks to generate devastating consequences for the poor. If the social safety net program can address market failures more directly, they may be able to help households break out of a persistent state of poverty.

For example, many transfer programs make the payments to households conditional on behaviors that encourage investment in health and education. Prominent examples include Mexico's Progresa/Oportunidades program and Brazil's Bolsa Família. More recently, more multi-faceted programs have arisen that explicitly aim to "graduate" households out of extreme poverty by providing households with a variety of 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.

We discuss the typical designs of different transfer programs, ranging from redistributive programs to those that aim to change the long-run prospects of the poor, and provide guidance for thinking about how to evaluate them. In particular, we discuss experimental methods to identify whether the programs are well-targeted to the poor, as well as whether the programs have the intended theoretical impacts. In doing so, we also provide a review of the current literature and discuss open questions.[1]

We pay particular attention to thinking about the theory behind different forms of social protection programs. Policymakers need to know not only that some program "works," but also why it works, i.e. the underlying mechanisms, because as economies advance and markets develop, program adaptations no doubt will be needed. Hence, the need is not only for evidence on what works, but for theory-linked evidence that can help provide lessons for future environments, societal and political pressures and market conditions.

II. The Varying Goals of Social Protection Programs

Social protection programs come in all shapes and forms. The appropriate program design will greatly depend upon the government or non-profit's primary goal. While an organization may have overlapping social goals, and the programs themselves may have a wide variety of impacts, we classify programs into three goal categories for ease of exposition.

[1] Summarizing a literature as broad and sizable as this poses a unique set of challenges. We have tried to focus on the key ideas in this area 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 in order to keep this chapter a reasonable length. We apologize in advance to those whose papers that we do not cover in detail.

The first type of program is one of *pure redistribution*. While these programs can differ in actual design, they share the common feature of first identifying the neediest families and then providing them with cash or in-kind transfers. Of course, pure redistribution programs could still have long-run growth impacts, e.g. if they are large enough to eliminate start-up costs for households to invest in particular forms of 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).[2] But, rather these types of programs primarily aim to limit poverty and hunger by ensuring that households attain a minimum living standard.

The second type of program also aims to redistribute, but then concurrently *improve the long-run outcomes of children*. This includes most conditional cash transfer programs (CCTs), which provide assistance to poor families conditional on children attending school and receiving vaccinations or other preventive health measures. These programs fulfill the goal of redistributing to poor households, while also improving future opportunities through increasing family investment in health and education.

Finally, we explore programs that aim to *generate a sustainable long-term income source* for poor households. These programs differ from pure redistributive programs in that that they include efforts to build long term income generating activities. Such programs include productive asset transfers, and more multi-faceted programs that also provide business or skills training, consumption support, health services and access to financial institutions. "Workfare" programs also share this aim, by providing transfers conditional on labor force training and/or participation. The design of these programs is often more complicated than pure redistribution programs. For example, programs must target households appropriate for the income generating activity, and programs must consider their interaction with product and labor markets.

[2] Notable examples that have found changes increases 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.

III. Redistributive Programs

In planning redistributive programs in developing countries, the first step is often deciding whom to target and how. Thus, in Section III.A., we first discuss the potential strengths and weaknesses of common targeting methodologies, as well as give the key factors that one must consider when planning RCTs on targeting methodologies. In Section III.B, we then discuss the particular form a program should take, e.g. should one provide cash or in-kind transfers? How big should the transfers be? How long should the transfers last? In doing so, we both recount what we know and do not know from the current experimental literature and also provide a guide for design of RCTs to evaluate these types of programs.

A. How should the Poor be Targeted?

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. 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.[3]

In order to fill this data gap, low-income country governments can conduct an income or consumption census. This too presents a challenge, 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, causal labor the next) or consumption (e.g. items purchased, the crops that one grows). Plus, there is nothing to stop someone from lying if they know that a cash prize is attached to their answers, since there is no formal mechanism in which to cross-check their data.

---

[3]In this section, we focus on targeting *poor* households, given that is this is the goal of redistributive programs. 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 would choose would depend upon whether you would like to redistribute scholarships to the poor, or 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 in Section 5, when we discuss programs that aim to change long-run outcomes.

As such, low-income countries tend to find alternative methods in 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 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? Thus, we will first start by outlining the four key categories of targeting methodologies and then talk about how experimental methods can be utilized to distinguish between varying features of these methods.

1. *Targeting Methods*

There are four key categories that encompass most targeting methodologies:

- Geographic: If the poor are concentrated in particular villages, districts, or regions, treating everyone in those areas may be an effective method to transfer resources to the poor (see, for example, Baker and Grosh, 1994; Elbers, et al 2007). This may be a particularly attractive form of targeting when the institutional capacity needed to collect individual information is low, as only aggregate information is needed, e.g. poverty mappings, rainfall data, etc. But, this 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 is a quick-and-dirty, poverty score card with just a few questions. Other times, a longer, more detailed asset and demographic survey is

[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.

[5] This is typically done with a nationally representative dataset 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.

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 that pass the proxy means test—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 ones in which everyone is allowed to apply, but in which some sort of "barrier" is put into place in order to reduce the probability that rich households access the program. Theoretically, there are different barriers that can generate this form of selection (Nichols and Zeckhauser, 1982), ranging from time costs, the quality of the good (Besley and Coate, 1991), means testing on arrival (Alatas, et al, 2015), and work requirements (Besley and Coate, 1992; Ravallion, M. 1991).[6] Done right, this can effectively screen out the rich (see, for example, Alatas, et al, 2015; Christan, 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, this also runs the risk that it may be hard in advance for governments to predict how many people will actually participate.

- Community based methods: Community members choose who in their locality are needy. 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

[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 IV, when we discuss programs that are geared at changing longer-run income or behavior.

this may make the program more politically popular, as people feel that the list is more in-line with their vision of who is deserving. The 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 doubles when the PMT is allowed to use its cross-village information.

The method and design should depend on the ultimate goals, context, institutional capacity, and targeting budget. For example, Alatas, et al (2012) shows that the choice of community methods versus PMT can depend on whether one 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) shows that aid allocations (for example, whether they reach the neediest or not) can depend on the type of community institutions that participate in targeting.

In practice, governments often mix and match methods depending on circumstance. For example, governments 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 PMT only in 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 methods to help determine the list of households that should 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, which are then given a PMT to determine actual eligibility (The Kenya CT-OVC Evaluation Team, 2012).

2. *Experimentally Testing Between Targeting Methods*

Are experimental methods important for testing across targeting approaches? Not necessarily. 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 method may prove attractive in some ways, it may miss out on the nuances of targeting that may be quite important in understanding the ultimate effectiveness of the methods. First, many of these targeting methods require effort on the part of citizens and the staff.[7] Thus, you would want conditions to be as real as possible, with at least a small amount of cash on the line for the households that are ultimately selected.[8] In terms of staff, the results may also differ considerably if you test out methods with a highly trained set of staff without stakes than if trying out the methods with typical staff in a real situation (Alatas, et al, 2013). The citizens might also behave differently during the process if they do not believe that the targeting list will actually be used to distribute resources. For example, individuals 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 have if there was 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.

Second, the choice of targeting method may affect both 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 outcomes. Targeted programs can be controversial, with one household receiving a transfer, while

[7] One obvious exception to this is geographic targeting, which does not rely on human behavior. If a different method was conducted in reality, 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 the geographic targeting in practice, it will not be able to feasible other aspects (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 any of the exercise very seriously.

their neighbors do not. If people believe a method is unfair and that it would produce a flawed list, they may be less likely to support the program, making it difficult to implement the program.[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.

Similarly, who is chosen may be different than who actually receives the transfer (Alatas, et al, 2013) and this may vary by targeting method. 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 conducting both methods in the same area to elicit a beneficiary list would possibly wrongly suggest that one should select the PMT. Instead, we would want to randomize the targeting methods to different areas to see how the method affects who ultimately receives the transfers.

The optimal design for a field experiment in the domain of targeting will depend on a number of factors, including what the methods 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 receives an intervention with that 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

[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.

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 the methods. Thus, in most cases, it is appropriate to randomize across a sensible geographic unit and thus 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 this case, a baseline is essential: a key outcome of any targeting experiment will be the 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 you being testing 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 error can be computed under the different methods.[10] To measure distortions along certain dimensions, it may be worthwhile to collect data on whether individuals are related to local leaders or political affiliations in the baseline. In the endline surveys, valuable data to collect could include who actually received the program, satisfaction levels, and 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, only need 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 sub-

[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.

randomizations: For example, Alatas, et al (2015) experimentally compare a PMT with self-targeting within Indonesia's conditional cash transfer program. Importantly, they experimentally vary the distance of the application site under self-targeting in order 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.

B. What Should Be Distributed?

Once the poor have been identified, there is a whole next set of questions: what should be given out? How much should be given? When should it be given and how often? To start thinking about these types of questions, it is useful to first define two main categories of unconditional, redistributive transfer programs:

- ***In-Kind Programs:*** These programs typically provide free or highly subsidized goods, to program recipients. They can be the direct provision of the goods, such as a food transfer program or a kerosene subsidy through local governments, NGO, or designated shops. Or, they could be vouchers that are constrained to particular types of goods, such as a food stamp program.[11] Examples include Indonesia's Rice Subsidy Program ("Raskin") and the Public Distribution Systems in both Bangladesh and India.
- **Unconditional Cash Transfer programs (UCT)**: While these may vary in terms of the specifics of how the program is run (i.e. amount, duration), they are typically characterized by households receiving a cash transfer with no strings attached. For example, 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

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

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.

In both types of programs, access is based merely on a targeting filter to identify a certain demographic group or income level, with no stipulations for how the money or goods should be used. These programs are "unconditional" in the sense that there are also no requirements on the specific behaviors that are required to access the funds, such as school or health clinic attendance. Both types of programs are common across the developing world, with Gentilini et al (2014) noting that 89 low-income countries have unconditional in-kind transfer programs and at least 119 countries have unconditional cash transfer programs.

Which is the best type of program? 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 often a series of arguments proposed in favor of in-kind subsidies (see Currie and Gahvari (2007) for an excellent review), which may explain their general persistence worldwide.

The most cited explanation is that of paternalism: people often hold a vision of the lazy, out-of-work husband coopting the cash and wasting it on alcohol, tobacco and entertainment, rather than making spending decisions that improve the family's living situation. Thus, that argument follows that an in-kind subsidy could reduce the husband's ability to do so, forcing redistribution within the household to women and children who typically have less household bargaining power. However, others argue that if the in-kind subsidy is infra-marginal—or 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 1996).

A second reason cited in support of in-kind programs is their self-targeting properties (e.g. Besley and Coate, 1991; Christan, 2014): by providing a good that the poor differentially value relative to the rich, the poor will apply and the rich will opt out. In practice, however, many transfer programs are independently targeted prior to distribution, shutting off a channel though which an in-kind program may generate these types of effects.[12] And, while the non-experimental evidence suggests that in-kind programs are better at selecting the poor (see for example Jacoby, 1997), there is little experimental evidence available that compares the magnitude of its targeting properties against different forms of means-tested cash programs.

A third potential reason to favor in kind transfers over cash transfers is that they might have different impacts on labor supply. 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 generate 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 Alzua, et al 2012), perhaps due to the long duration of benefits and uncertain processes for re-certification 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 (e.g. Gertler, et al, 2006). Understanding the relative impacts of in-kind versus cash transfer programs on labor market participation remains an area for active research.

[12] A notable exception is India's NREGA program, which is a universal program that requires physical work (e.g. on construction projects) to access payments. In this case, who chooses to engage in work may be a function of their initial poverty status.

A number of randomized control trials have been conducted to evaluate the impact of unconditional transfer programs. In these types of studies, households are first targeted in the sample area and then potential beneficiaries are randomly assigned to treatment and control to assess 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 the Malawi Social Cash Transfer Scheme (Covarrubias, et al , 2012).[13]

This basic RCT comparing the effect of a specific program against control areas that do not receive any assistance is important for understanding general program effects. However, we often care not just whether a program has an effect, but how it would fare against different program types (e.g. in-kind versus cash). In this case, *multiple treatment arms* that compare different programs can be useful. For example, there have been a number of nice studies of late that have compared in-kind and cash transfer programs. For example, the Mexican government 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 (3) a control group that did not receive transfers.[14] While the in-kind program distorted consumption of some individual types of food (Cunha, 2012), 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). Similarly, 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

---

[13] 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 and Araujo, 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 below.

[14] Note that a particular challenging 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 percent 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, 2012 for a discussion of how to make them ex-post equivalent.

regardless of the mechanism since voucher households simply bought food that is relatively easy to sell (e.g. salt).

In contrast, other studies highlight the tradeoffs between programs that may satisfy society's paternalistic goals (e.g. greater food consumption) versus other important social goals. Hidrobo, et al (2013) 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. 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, one can think about how to model the tradeoff between the additional utility households receive from spending as they choose relative to society's utility from the shift in food consumption.

Importantly, all of these papers highlight that the transfer programs can have many different impacts on a household and that these impacts may differ based on the specific design of a program and the initial conditions of an area—e.g. whether a particular item of food is intra-marginal or not. Thus, the use of multiple treatment arms can also help us isolate different theories of how a particular type of transfer works in practice, which can be useful in generating results that can be more easily generalized to other contexts. For example, if markets functioned perfectly, households would spend money in the same way regardless if they received small, frequent amounts of cash or one big transfer. However, if there are savings or credit constraints, theory may predict very different spending patterns. Thus, Haushofer and Shapiro (2014) use multiple treatment arms to not only test for an effect of the transfer program, but to test this particular theory: they randomize whether households get a lump-sum cash transfer or smaller monthly cash transfers. Indeed, they find evidence of market imperfections, as the lump-sum gets spent on durables, while the monthly sum gets spent to improve food security.

There are still many open questions in which the use of more sophisticated multiple treatments arms could be useful. For example, we still know relatively little about labor market effects of transfer programs in developing countries. One can test whether having a sharp cutoff for eligibility versus different forms of a gradual decline in eligibility cutoffs would have different types of labor market effects, especially under the different levels of observability of income and formality of employment observed in different parts of the developing world.

Practically, once the research questions and treatments have been identified, there are a number of specific issues that need to be considered when designing an RCT in this space. First, at what level should the randomization take place? Transfers are generally an individually based program, and so it is tempting to randomize at the individual-level to maximize statistical power. For example, Schady and Araujo (2008) do exactly this. However, Angelucci and de Giorgi (2009), among others, show that there may be spillovers from eligible to ineligible households within a village. This implies that the randomization likely needs to be done at a higher level, at the village-level or sub-district level depending upon what types of spillovers one might expect.[15]

There are two other aspects about the randomization to keep in mind: First, if randomization is at a group level, it is important 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 program impacts; the Malawi Social Cash Transfer Scheme had only 8 clusters and did not fully account for the grouped nature of the data (Covarrubias, et al 2012).[16] Thus, it is important to 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 whether you will have enough statistical power to estimate them.

---

[15] As we talk about below, one might even want to design the study to capture different types of spillovers and general equilibrium effects.

[16] 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).

Second, who are the beneficiaries in the control group? Suppose 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 know who were the hypothetical beneficiaries in the control group (see, for example, Covarrubias, et al 2012, who had this problem). In this case, it would be possible to estimate the impact on the entire village as a whole, but not easily be able to estimate the impact on just the beneficiaries.[17] Thus, it is important, when possible, to use the same targeting methods in the treatment and control group, even if the control group is not receiving the program.

The second key evaluation design questions revolve around the data collection. What data should be collected and when? Starting with what data to collect, these types of programs are meant to redistribute to the poor. But, in many cases, due to weaker institutional structures, corruption, or imperfect information on their entitlements, households do not receive the full transfer. So, a first set of survey questions should be focused on whether households actually received the transfer and if it had an effect on consumption. The rest of the survey questions should arise from the different theoretical impacts that one may believe are important in understand program impacts or due to the particular context. For example, in comparing a food transfer program with a cash one, 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. But, if the goal of a food transfer program relative to cash is less about changing consumption patterns, but more about exploiting self-targeting properties, then initial poverty and income status of households is key to measure.

Note two features of the data collection process. First, as Schady and Araujo (2008) point out, people may worry about providing accurate information on surveys if they believe that the surveys are connected to re-verification for the program. In which case,

[17] Of course, this is not a problem if the program is geographically targeted and everyone in all sample villages is eligible.

one may want to collect variables that are easier to verify: assets that one can observe, vaccination records on a card, body mass index (BMI) and other anthropometric variables to assess the health status of children, etc. Second, as transfer programs may have a multitude of different outcomes, one worry is that in collecting many different variables, we would find some significant impacts just by chance. Thus, one might want to pre-specify key hypothesis and outcome variables in advance (Miguel et al, 2014).

When should data collection take place? A first 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 post-intervention of the two groups is an unbiased estimate of the true impact. There are three exceptions: (1) if the targeting process is of interest, there is value in collecting baseline data on household income or consumption levels (2) if the pilot program is constrained in terms of how many areas are treated by budget, collecting and controlling for baseline characteristics of households in analysis can improve statistical power or (3) if the question of treatment effects on particular sub-groups is of importance (e.g. the very poor, families with children who are less likely to attend school, etc.).

A second question is when to conduct the 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 and to assess spending patterns. However, as we discuss below, if the longer-run impacts of transfer programs are of interest, one may also want to do additional surveys sometime after a household's is no longer enrolled in the program.

IV. Conditional Transfers

Redistribution policies often include conditions, e.g., conditional cash transfers (CCTs). CCT programs have become common in Latin America and have begun to spread to other parts of the world, with CCT programs in more than 52 developing countries as of 2013 (Gentilini et al, 2014, Fiszbein and Schady 2009; Saavedra and Garcia 2012). Typical CCT programs employ proxy-tests for eligibility and then make regular cash transfers to households conditional on household compliance with certain long-term investments, such as children attending school or visiting health clinics. The programs have generated important impacts, but are also quite expensive. If ignoring the redistribution motivation (which would be wrong to do), they rarely pass cost-benefit tests for just the behavior change targeted by the conditionality. The logic behind such programs is two-fold: (a) a distributive element which shares the motivation, program and targeting design issues discussed in the Redistribution section, (b) a conditionality which typically aims to shift household choices towards long term investments in health or education.

Since we just discussed several redistribution program design and evaluation issues that also apply to conditional transfer programs, here we focus strictly on the conditionality.

A. Condition (CCT) versus No Condition (UCT)

Evaluations that test a conditional transfer versus no program have difficulty 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 as they require budget and staff to enforce, and they may also generate selection effects on who chooses to participate. To test the effects of the conditionality, a simple experimental design would have one treatment with a conditional cash transfer, a second treatment with unconditional cash transfers, 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 changed behavior simply through its income effect.

This is the approach taken by Baird, McIntosh and Özler (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 test scores of cognitive ability, mathematics, and English reading comprehension increased for the CCT arm, but not the UCT arm. However, the UCT arm had significantly lower marriage and pregnancy rates. This initially may feel like a counter-intuitive 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, Dupas, and Kremer 2010; Ferre 2009; Osili and Long 2008; Ozier 2011), the UCT will lead to, in net, lower marriage and pregnancy rates, and this is what the authors found in the Malawi study. This reinforces the need for careful data collection on ancillary outcomes–e.g. in this case, marriage and pregnancy—in order to understand the full set of mechanisms through which a program works so that policy-makers can better understand the tradeoffs that they would make by implementing one program over another.

Further insights on the tradeoff between a UCT and a 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, Karlan, and Yin 2006; Bryan, Karlan, and Nelson 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 a CCT on school attendance, and also tested a sub-treatment 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. This sheds important insights into the underlying mechanisms of the CCT, and suggests that the conditionality in this context was simply a tool for parents to better monitor children, and thus better interventions to improve monitoring and communications between schools and parents may be a better solution than the complicated nature of CCTs over UCTs.

B. What should on the Conditions be?

When choosing to implement a conditional transfer program, it is essential to determine which conditions should be imposed. For example, do you want to condition on schooling outcomes? 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. Once these conditions are determined, 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 re-enrolls 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 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 (2014) 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 above, applying conditions may lead to a tradeoff between the goals of the conditios 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? Thus, one can imagine extensions of the CCT literature where you randomize not only whether the program has conditions, but to randomize the types of conditions in different areas—measuring not only the effect on recipients, but also on who applies for the programs and how well the implementers can realistically enforce them.

C. Enforcement of the Conditionality

The enforcement of the conditionality is critical to examine. Without the enforcement, the program is perhaps theoretically equivalent to an unconditional cash transfer. 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 appropriately analyzes the results as if the program is an unconditional cash transfer (Fernald and Hidrobo 2011; Paxson and Schady 2010). Naturally, a CCT program which 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. This is an area worthy of further research, but it is not obvious this is 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 that merely randomizing enforcement in one setting does not translate to a setting in which enforcement simply was not possible for political or social reasons. This could be driven by expectations of enforcement, social norms which act as omitted variables (which both drive the level of enforcement and the treatment effect of the program), or likely accompanying policies in the two settings which 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 labelling a transfer as an "education support program." Benhassine et al (2015) examines this question by designing a two-prong experiment: conditional cash transfer versus labelled cash transfer (those two prongs were crossed with a household structure test, providing the program to mothers versus fathers). In order to understand the underlying mechanisms, the study collected data on process changes and also used a multi-armed experimental design. For example, to understand if the conditionality (or labelling) 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 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 (which thus draws a similar conclusion on mechanisms as the Brazilian study above (Bursztyn and Coffman 2012). The multi-arm 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.

V. Transfers to Increase Long-Term Income

Improving long-term income for the poor poses a more complicated challenge, in that one likely needs to address underlying mechanisms that are creating the poverty traps. If the underlying market frictions are beyond credit and savings market constraints, then the solution will require more than mere redistribution. If redistribution alone is employed, it may provide important short-run benefits, but act as more of a band aid on immediate symptoms and not help individuals achieve a sustained increase in income.

For the past thirty years, microcredit has been a leading development policy in the fight to reduce poverty. Unfortunately, seven recent randomized trials have shown that microcredit, while it does provide some important benefits, does not improve long term income, on average, for participants (for a review, see Banerjee, Karlan, and Zinman 2015; the seven randomized trials are Angelucci, Karlan, and Zinman 2015; Attanasio et al. 2015; Augsburg et al. 2015; Banerjee et al. 2015; Crépon et al. 2015; Karlan and Zinman 2011; Tarozzi, Desai, and Johnson 2015).

When it comes to combating poverty, the whole may be greater than the sum of the parts. We typically see programs tackling one problem at a time. Poverty alleviation policy, as with many government programs, often operates in silos. One silo, described above and typically managed under the umbrella of "social protection," focus on redistribution policies through either conditional or unconditional cash transfers. A second silo, often managed under the ministries of trade or agriculture, focus on livelihood support, such as the transfer or a productive asset or agricultural input, alongside some training. A third silo, financial inclusion, is often pushed by the ministry of finance or directly through subsidized or market-driven 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 try to improve their living standard.

The “Graduation” approach is an integrated, multi-faceted program with livelihood promotion at its core, and with an aim 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, 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 multi-faceted program was “sufficient” to increase long-term income, where long-term is defined as three years after the productive asset transfer. The results from the pooled analysis across all six countries found that the program led to sustainable and significant impacts in consumption, food security, asset wealth, savings and revenues [insert results]. This “sufficient” argument, however, while important for policy, does not provide insight into the underlying theoretical mechanisms.

These results are promising, in that they show that a sufficient set of interventions is capable of alleviating poverty sustainably. They should whet the appetite, both for a more theoretically grounded understanding of which markets failures exactly 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 was 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 2x2x2x2 = 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 (but not necessarily the stock) must be observed. 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, 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, and the way forward is clearly going to be the development of a mosaic, rather than any one definitive study that both tests each component and also includes sufficient contextual and market variations that it can help set policy for a myriad of countries and populations. For example, in a post-conflict setting in Uganda, a NGO-led program provided youth groups with training and cash (US$150) towards non-agricultural self-employment activity and found a 57% increase in business assets, 17% increase in work hours, and 38% increase in earnings four years after the cash grants (Blattman, Fiala, and Martinez 2014). This program differs from the above mentioned Graduation programs in three potentially important and illuminating dimensions: post-conflict versus non post-conflict, 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 variation, suggests 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. 2014). These programs, as with above Uganda program, differ from the Graduation studies in that they do not include savings, health and life coaching components, and are focused on enterprise development (rather than animal husbandry, the dominant livelihood in the Graduation studies).

Thus, the initial studies above 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, while the Graduation studies cited above measured impacts three years after the assets were transferred, the implementing organizations continued to engage with households two years after the asset transfers. If the household visits were 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 one year after the household visits ceased.

Second, as some of the studies above 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 above 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 [**add psych cites**]. 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, a combination of experimental design and data can help shed insight into such effects. The first task is to be more specific, as general equilibrium effects encompass a wide variety of indirect effects: for example, price of transferred assets; spillovers from explicit sharing of granted resources; and increased economic activity from increasing the wealth of the poor. A typical experimental design would either randomize across and within villages (this requires assuming that the village is the boundary for generating general equilibrium effects), or for some issues, examining spillovers to non-participants in treatment versus control. The latter was the approach taken in Mexico, in which researchers found that conditional cash transfer recipients lent money to non-participants (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 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.

VI. Ideas Only Go So Far: Implementation Matters Too

A transfer program may look like a winner on paper, but 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 anti-poverty programs, it is also important to think about whether, theoretically, a particular aspect of the program implementation is likely to be particularly vulnerable to implementation problems. There are two types of variations one can think about: (1) evaluations that vary the underlying structure of the program (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 especially 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 should one make the transfers—may matter a lot, affecting the level of leakages and corruption, targeting, the costs for beneficiaries to access program, and potentially how beneficiaries spend their entitlements. For example, there is extensive work, in general, exploring how officials' incentives affect their work output, 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 Bo, et al 2013; Ashraf, et al 2014; Hanna and Wang, 2014), but less specifically on how changing who is selected to implement transfer programs affects outcomes. For future research, one can imagine varying the salary structure both for current workers and in the recruitment of new workers to look at how it affects targeting and delivery, as well as varying what officials are incentivized on (e.g. whether they are specifically incentivized on enforcing CCT conditions).

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 could also bid for the right to run the program. They find that the bidding reduces the price-market 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 re-evaluated. 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 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, to net out potential effects of a cell phone from mobile money. Mobile money reduced the non-profit's distribution costs and also reduced the households' costs to pick up their entitlement. This second feature of mobile money many be particularly important if we believe high transaction costs induce beneficiary households "leave money" on the table (Currie and Gahvari, 2007). 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) aims to also get at some of these types of questions. They evaluated the impact of biometrically-authenticated payments infrastructure ("Smartcards") on beneficiaries of employment (NREGS) 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 sub-districts, so the authors randomized which sub-districts were converted first. Following the introduction of the program, not only did the time it took beneficiaries to collect a payment fall, but the delay in receiving the payment was reduced by almost 30%. The ease of payment induced households to work more. Households, thus, earned more, while payments to officials remained the same—hence, leakages fell quite dramatically.

Experimentally testing **complementary programs** that are layered on top of existing programs can also be important in improving delivery. These programs do not require changing the existing program rules or functioning, but instead provide additional services to help citizens better access programs. For example, one could test how information on eligibility and program rules helps citizens better access programs. Ravallion, et al (2013) 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 reduces leakages in a subsidized rice program, and that making the card information public within the village has even larger impacts. Alternatively, one can imagine experiments designed to test how providing households with direct help with their paperwork when applying affects who enrolls.

One final point on implementation relates the broader debate about conducting field experiments in settings related to how they would scale. 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; Hanna and Dhaliwal, 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 NGO to that government. NGOs are not a monolithic set of institutions, whereas neither are governments. Thus it may be the wrong to ask whether 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 and behavioral responses. 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 as stated in the past.

In terms of implementation efficiency, one can also imagine 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 implemented perfectly, citizens would be indifferent between cash and an in-kind transfer. 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 across the programs. Again, this can be a difference between government and an NGO, but speaks to larger differences in what are the relative implementation abilities of different organizations and how can one improve upon their weaknesses.

VII. Future Work and Open Questions

Since the innovative and instrumental randomized evaluation of Progresa in Mexico, there has been a burst of important and exciting work 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 help provide them with growth opportunities.

So, then the question becomes, where should we focus our research efforts next? Within this chapter, we have tried highlight gaps in our understanding of the design of anti-poverty programs, ranging from how to structure conditions or disentangling the effects of more multi-faceted transfer programs. We end this chapter by highlighting several areas within this space that we believe are *especially* important for future research:

A. Long-run Impact

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.

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. For example, 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 re-certification 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?

B. Interaction of Demand and Supply Side

A vital question is how transfer programs work across different contexts, particularly in presence of supply side constraints. 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. Similarly, the Graduation Program had large effects in most countries, except Honduras. Thus, further research is necessary to understand why the effects may have varied, from the initial difference in poverty status to different institutional features of these areas.

More broadly, there are lots of unanswered questions about the interaction of transfer programs with existing conditions: for example, how does school quality, or health care availability, affect the adherence to CCT conditions? Does food or other in-kind transfers work better than cash in areas with more limited food supplies? Do transfers work augment 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. To test these kinds of questions, 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.

C. General Equilibrium Effects

With relatively large sums being distributed, anti-poverty program may have broader effects than one initially expects, with these effects being potentially quite large. These effects can take various forms, affecting insurance and lending markets within villages (Angelucci and de Giorgi, 2009) to affecting natural resource demand (Hanna and Oliva, 2015; Gertler, et al 2013) to labor markets (Muralidharan, et al 2015).

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 positive peer effects on the schooling outcomes of ineligible children in Mexico's Progressa (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.

[insert Graduation spillover results here]

While there have been a few studies, including those discussed above, 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 different types of general equilibrium effects that may be possible, this is a case where we are particularly worried about multiple hypothesis testing. Careful theory of what we would expect under different types of programs and conditions, coupled with pre-specified hypotheses, may be important in measuring these effects.

Identifying spillover effects should be built into the research plan when plausible and viable. Three basic approaches have been employed: (a) 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), (b) 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 Giorgio 2009), and (c) through data collection on process changes: for example, collecting specific data on informal transfers, credit and savings could identify behaviors that indicate the presence general equilibrium effects.

D. Social Insurance

Much of the research on anti-poverty programs in developing countries to date has been focused on transfer programs. However, as developing countries transition, we expect the importance of other forms of anti-poverty programs to also grow. For example, given shifting demographics and the risk of poverty for the elderly in many developing countries (World Population Aging Report, 2013), non-contributory pension programs exist in many countries—such as Brazil, South Africa, India, etc—as a tool to combat poverty for this particularly vulnerable group. 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, whether the pensions are targeted, etc. 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.

Similarly, as countries grow, employment is shifting from agricultural work to more formal forms of employment. As such, identifying who is facing a temporary shock becomes easier, and so we would expect a shift from programs that provide a generally poor household with transfers for X number of months or years to those that aim to provide assistance to households at the time they experience a shock. For example, richer developing countries (e.g. Brazil, Egypt) tend to have unemployment insurance programs that mimic those seen in more developed nations. Thus, the need for more research and understanding of how these programs can be made to “best” work in developing country settings will become increasing important in the coming decade.

WORKS CITED

Aker, Jenny. 2014. "Comparing Cash and Voucher Transfers in a Humanitarian Context: Evidence from the Democratic Republic of Congo." Center for Global Development Working Paper No. 320.

Aker, Jenny C., Rachid Boumnijel, Amanda McClelland, and Niall Tierney. 2014. "Payment Mechanisms and Anti-Poverty Programs: Evidence from a Mobile Money Cash Transfer Experiment in Niger." Center for Global Development Working Paper No. 268.

Akresh, Richard, Damien de Walque, and Harounan Kazianga. 2013. "Cash Transfers and Child Schooling: Evidence from a Randomized Evaluation of the Role of Conditionality." World Bank Policy Research Working Paper No. 6340.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review*. 102(4): 1206-1240.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Ben Olken, Ririn Purnamasari, and Matt Wai-Poi. 2013. "Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia." NBER Working Paper No. 18798.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2015. "Self-Targeting: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, forthcoming.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Ben Olken, Matthew Wai-Poi, and Ririn Purnamasari. 2014. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." International Initiative for Impact Evaluation, Impact Evaluation Report No. 12.

Angelucci, Manuela, Giacomo de Giorgi, Marcos A. Rangel, and Imran Rasul. 2010. "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment." *Journal of Public Economics*. 94(3-4): 197-221.

Angelucci, Manuela and Giacomo de Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review*. 99(1): 486-508.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 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, London School of Economics Working Paper 54.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics*. 126(4): 1709-1753.

Baker, Judy L. and Margaret E. Grosh. 1994. "Poverty Reduction Through Geographic Targeting: How Well Does It Work?" *World Development*. 22(7): 983-995.

Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto. 2014. "Information is Power: Identification Cards and Food Subsidy Programs in Indonesia." Working Paper.

Bardhan, Pranab K. and Dilip Mookherjee. 2000. "Capture and Governance at Local and National Levels." *American Economic Review*. 90(2): 135-139.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics*. 3(2): 167-195.

Barrera-Osorio, Felipe and Deon Filmer. 2013. "Incentivizing Schooling for Learning: Evidence on the Impact of Alternative Targeting Approaches." World Bank Policy Research Working Paper No. 6541.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2014. "Self-Selection into Credit Markets: Evidence from Agriculture in Mali." Working Paper.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2014. "Self-Selection into Credit Markets: Evidence from Agriculture in Mali." National Bureau of Economic Research Working Paper No. 20387.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2013. "Do Elected Councils Improve Governance? Experimental Evidence on Local Institutions in Afghanistan." MIT Political Science Department Research Paper No. 2013–24.

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

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

Besley, Timothy, and Stephen Coate. 1991. "Public Provision of Private Goods and the Redistribution of Income." *American Economic Review*. 81(4): 979-984.

Besley, Timothy and Stephen Coate. 1992. "Workfare Versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs." *American Economic Review*. 82(1): 249–61.

Bobonis, Gustavo J. and Frederico Finan. 2009. "Neighborhood Peer Effects in Secondary School Enrollment Decisions." *Review of Economics and Statistics*. 91(4): 695-716.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. "Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education." Center for Global Development Working Paper No. 321.

Chen, Shaohua, Martin Ravallion, and Youjuan Wang. 2006. *Di Bao: A Guaranteed Minimum Income in China's Cities?* Vol. 3805. Washington, DC: World Bank Publications.

Christian, Paul. 2014. "The Distributional Consequences of Group Procurement: Evidence from a Randomized Trial of a Food Security Program in Rural India." Job Market Paper.

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

Cohen, Jessica and Pascaline Dupas. 2010. "Free Distribution or Cost-sharing? Evidence from a Randomized Malaria Prevention Experiment." *The Quarterly Journal of Economics.* 125(1): 1-45.

Cohen, Jessica, and Pascaline Dupas. 2007. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." Brookings Global Economy and Development Working Paper No. 11.

Covarrubias, Katia, Benjamin Davis, and Paul Winters. 2012. "From Protection to Production: Productive Impacts of the Malawi Social Cash Transfer Scheme. *Journal of Development Effectiveness*. 4(1): 50-77.

Cunha, Jesse M. 2012. "Testing Paternalism: Cash vs. In-kind Transfers." *American Economic Journal: Applied Economics*. 6(2): 195-230.

Cunha, Jesse M., Giacomo de Giorgi, and Seema Jayachandran. 2013. "The Price Effects of Cash Versus In-Kind Transfers." NBER Working Paper No. 17456.

Currie, Janet, and Firouz Gahvari. 2007. "Transfers in Cash and in Kind: Theory Meets the Data." National Bureau of Economic Research Working Paper No. 13557.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *The Quarterly Journal of Economics.* 128(3): 1169-1218.

Dhaliwal, Iqbal, Rema Hanna. 2014. "Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India." NBER Working Paper No. 20482.

Duflo, Esther, Rachel Gelnnerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*. Volume 4, eds. T. Paul Schultz and John A. Strauss, 3895-3962. Amsterdam: Elsevier.

Dupas, Pascaline, Vivian Hoffmann, Michael Kremer, and Alix Peterson Zwane. 2013. "Micro-Ordeals, Targeting, and Habit Formation." Working Paper.

Edmonds, Eric V. and Norbert Schady. 2012. "Poverty Alleviation and Child Labor." *American Economic Journal: Economic Policy*. 4(4): 100-124.

Elbers, Chris, Tomoki Fujii, Peter Lanjouw, Berk Özler, and Wesley Yin. 2007. "Poverty Alleviation Through Geographic Targeting: How Much Does Disaggregation Help?" *Journal of Development Economics*. 83(1): 198-213.

Fernald, Lia C.H. and Melissa Hidrobo. 2011. "Effect of Ecuador's Cash Transfer Program (*Bono de Desarrollo Humano*) on Child Development in Infants and Toddlers: A Randomized Effectiveness Trial." *Social Science & Medicine*. 72(9): 1437-1446.

Galiani, Sebastian and Patrick J. McEwan. 2013. "The Heterogeneous Impact of Conditional Cash Transfers." *Journal of Public Economics*. 103(C): 85-96.

Gentilini, Ugo, Maddalena Honorati and Rsulan Yemtsov. 2014. *The State of Social Safety Nets 2014*. Washington, DC: World Bank Group.

Gertler, Paul J., Sebastian W. Martinez, and Marta Rubio-Codina. 2012. "Investing Cash Transfers to Raise Long-Term Living Standards." *American Economic Journal: Applied Economics*. 4(1): 164-92.

Gertler, Paul, Orie Shelef, Catherine Wolfram, and Alan Fuchs. 2013. "The Demand for Energy-Using Assets Among the World's Rising Middle Classes." Energy Institute @ Haas Working Paper.

Glewwe, Paul, Albert Park, and Meng Zhao. 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 No. WP12-2.

Hanna, Rema and Paulina Oliva. 2015. "Moving Up the Energy Ladder: The Effect of a Permanent Increase in Assets on Fuel Consumption Choices in India." *American Economic Review Papers and Proceedings***,** forthcoming.

Hanna, Rema, and Shing-Yi Wang. 2014. "Dishonesty and Selection into Public Service: Evidence from India." NBER Working Paper #19649, submitted.

Haushofer, Johannes and Jeremy Shapiro. 2014. "Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya." Working Paper.

Jessee, Cassandra, Lean Prencipe, Dan Sherman, Alefa Banda, Liseteli Ndiyoi, Nathan Tembo, Silvio Daidone, Joshua Dewbre, Mario Gonzales-Flores, Ed Taylor, Karen Thome, and Amber Peterman. American Institutes for Research. 2013. "Zambia's Child Grant Program: 24-Month Impact Report."

Kazianga, Harounan, Damien de Walque, and Harold Alderman. 2012. "Educational and Child Labor Impacts of Two Food-for-Education Schemes: Evidence from a Randomized Trial in Rural Burkina Faso." *Journal of African Economies*. 21(5): 723-760.

Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz, and Lisa Sanbonmatsu. 2004. "Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment." KSG Working Paper No. RWP04-035.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics*. 91(3): 437-456.

Lalive, Rafael and M. Alejandra Cattaneo. 2009. "Social Interactions and Schooling Decisions." *Review of Economics and Statistics*. 91(3): 457-477.

Leroy, Jef L., Paola Gadsden, Teresa González de Cossío, and Paul Gertler. 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." *Journal of Nutrition*. 143(3):378-383.

Ma, Xiaochen, Sean Sylvia, Matthew Boswell, and Scott Rozelle. 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.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, M. Van der Laan. 2014. "Promoting Transparency in Social Science Research." *Science.* 343 (6166): 30.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2014. "Building State Capacity: Evidence from Biometric Smartcards in India." NBER Working Paper No. 19999.

Nichols, Albert L. and Richard J. Zeckhauser. 1982. "Targeting Transfers Through Restrictions on Recipients." *American Economic Review*. 72(2): 372-377.

Paxson, Christina and Norbert Schady. 2010. "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador." *Economic Development and Cultural Change*. 59(1): 197-229.

Ravallion, Martin. 1991. "Reaching the Rural Poor Through Public Employment: Arguments, Evidence, and Lessons from South Asia." *World Bank Research Observer*. 6(2): 153–175.

Ravallion, Martin, Dominique van de Walle, Puja Dutta, and Rinku Murgai. 2013. "Testing Information Constrains on India's Largest Antipoverty Programs." World Bank Policy Research Working Paper No. 6598.

Rofman, Rafael; Apella, Ignacio; Vezza, Evelyn. 2015. *Beyond Contributory Pensions: Fourteen Experiences with Coverage Expansion in Latin America.* Washington, DC: World Bank Publications.

Schady, Norbert and Maria Caridad Araujo. 2008. “Cash Transfers, Conditions, and School Enrollment in Ecuador.” *Economia: Journal of the Latin American and Caribbean Economic Association*. 8(2): 43-77.

Seabright, Paul**.** 1996. “Accountability and Decentralisation in Government: An Incomplete Contracts Model.” *European Economic Review.* 40 (1): 61–89.

Schultz, T. Paul. 2004. “School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program.” *Journal of Development Economics*. 74(1): 199-250.

Skoufias, Emmanuel, Mishal Unar, and Teresa Gonzalez-Cossio. 2013. “The Poverty Impacts of Cash and In-Kind Transfers: Experimental Evidence from Rural Mexico.” *Journal of Development Effectiveness*. 5(4): 401-429.

Southworth, Herman M. 1945. "The Economics of Public Measures to Subsidize Food Consumption." *Journal of Farm Economics.* 27(1): 38–66.

Vermeersch, Christel and Michael Kremer. 2004. “School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation.” World Bank Policy Research Working Paper No. 3523.

The Kenya CT-OVC Evaluation Team. 2012. “The Impact of Kenya's Cash Transfer for Orphans and Vulnerable Children on Human Capital.” *Journal of Development Effectiveness*. 4(1): 38-49.

Angelucci, Manuela, and Giacomo De Giorgi. 2009. “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption?” *American Economic Review* 99 (1): 486–508. doi:10.1257/aer.99.1.486.

Angelucci, Manuela, Dean Karlan, and Jonathan Zinman. 2015. “Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco.” *American Economic Journal: Applied Economics* 7 (1): 151–82. doi:10.1257/app.20130537.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. “Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines.” *Quarterly Journal of Economics* 121 (2): 673–97.

Attanasio, Orazio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harmgart. 2015. “The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia.” *American Economic Journal: Applied Economics* 7 (1): 90–122. doi:10.1257/app.20130489.

Augsburg, Britta, Ralph De Haas, Heike Harmgart, and Costas Meghir. 2015. “The Impacts of Microcredit: Evidence from Bosnia and Herzegovina.” *American Economic Journal: Applied Economics* 7 (1): 183–203. doi:10.1257/app.20130272.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. “Cash or Condition? Evidence from a Cash Transfer Experiment.” *The Quarterly Journal of Economics* 126 (4): 1709–53. doi:10.1093/qje/qjr032.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. “The Miracle of Microfinance? Evidence from a Randomized Evaluation.” *American Economic Journal: Applied Economics* 7 (1): 22–53. doi:10.1257/app.20130533.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015. “Six Randomized Evaluations of Microcredit: Introduction and Further Steps.” *American Economic Journal: Applied Economics* 7 (1): 1–21. doi:10.1257/app.20140287.

Banerjee, Abhijit, and Andrew Newman. 1993. “Occupational Choice and the Process of Development.” *Journal of Political Economy* 101: 274–98.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2015. “Turning a Shove into a Nudge? A ‘Labeled Cash Transfer’ for Education.” *American Economic Journal: Economic Policy, Forthcoming*.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. “Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda.” *The Quarterly Journal of Economics* 129 (2): 697–752.

Blattman, Christopher, Eric Green, Jeannie Annan, and Julian Jamison. 2014. “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.

Bryan, Gharad, Dean Karlan, and Scott Nelson. 2010. “Commitment Devices.” *Annual Review of Economics* 2 (1): 671–98. doi:10.1146/annurev.economics.102308.124324.

Bursztyn, Leonardo, and Lucas C. Coffman. 2012. “The Schooling Decision: Family Preferences, Intergenerational Conflict, and Moral Hazard in the Brazilian Favelas.” *Journal of Political Economy* 120 (3): 359–97. doi:10.1086/666746.

Chetty, Raj, and Adam Looney. 2006. “Consumption Smoothing and the Welfare Consequences of Social Insurance in Developing Economies.” *Journal of Public Economics* 90 (12): 2351–56. doi:10.1016/j.jpubeco.2006.07.002.

Crépon, Bruno, Florencia Devoto, Esther Duflo, and William Pariente. 2015. “Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco.” *American Economic Journal: Applied Economics* 7 (1): 123–50. doi:10.1257/app.20130535.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2010. “Education and Fertility: Experimental Evidence from Kenya.” *Working Paper*.

Fernald, Lia C.H., and Melissa Hidrobo. 2011. “Effect of Ecuador’s Cash Transfer Program (Bono de Desarrollo Humano) on Child Development in Infants and Toddlers: A Randomized Effectiveness Trial.” *Social Science & Medicine* 72 (9): 1437–46. doi:10.1016/j.socscimed.2011.03.005.

Ferre, Celine. 2009. *Age at First Child: Does Education Delay Fertility Timing? The Case of Kenya*. SSRN Scholarly Paper ID 1344718. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=1344718.

Fiszbein, Ariel, and Norbert Rüdiger Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. A World Bank Policy Research Report 47603. Washington D.C: World Bank.

Karlan, Dean, and Jonathan Zinman. 2011. “Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation.” *Science* 332 (6035): 1278–84. doi:10.1126/science.1200138.

McKenzie, David, and Christopher Woodruff. 2006. “Do Entry Costs Provide an Empirical Basis for Poverty Traps? Evidence from Mexican Microenterprises.” *Economic Development and Cultural Change* 55 (1): 3–42. doi:10.1086/505725.

Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115: 200–249.

Osili, Una Okonkwo, and Bridget Terry Long. 2008. "Does Female Schooling Reduce Fertility? Evidence from Nigeria." *Journal of Development Economics* 87 (1): 57–75. doi:10.1016/j.jdeveco.2007.10.003.

Ozier, Owen. 2011. "The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis." *Working Paper*. http://economics.ozier.com/owen/papers/ozier_JMP_20110117.pdf.

Paxson, Christina, and Norbert Schady. 2010. "Does Money Matter? The Effects of Cash Transfers on Child Development in Rural Ecuador." *Economic Development and Cultural Change* 59 (1): 197–229.

Saavedra, Juan, and Sandra Garcia. 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.

Tarozzi, Alessandro, Jaikishan Desai, and Kristin Johnson. 2015. "The Impacts of Microcredit: Evidence from Ethiopia." *American Economic Journal: Applied Economics* 7 (1): 54–89. doi:10.1257/app.20130475.