



When Are Field Experiments with Individual Assignment Too Risky?

**Lessons from a Center-Based
Child Care Study in Guatemala**

Sarah D. Humpage

**Inter-American
Development Bank**

Department of Research
and Chief Economist

TECHNICAL NOTE

No. IDB-TN-469

November 2012

When Are Field Experiments with Individual Assignment Too Risky?

Lessons from a Center-Based Child Care Study in Guatemala

Sarah D. Humpage



Inter-American Development Bank

2012

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Humpage, Sarah D.

When are field experiments with individual assignment too risky? lessons from a center-based child care study in Guatemala / Sarah D. Humpage.

p. cm. (IDB Technical Note; 469)

Includes bibliographical references.

1. Social sciences—Research—Methodology. 2. Early childhood education—Guatemala. I. Inter-American Development Bank. Research Dept. II. Title. III. Series.

<http://www.iadb.org>

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

The unauthorized commercial use of Bank documents is prohibited and may be punishable under the Bank's policies and/or applicable laws.

Copyright © 2012 Inter-American Development Bank. All rights reserved; may be freely reproduced for any non-commercial purpose.

Abstract*

Randomized controlled trials, prized for generating unbiased estimates of treatment effects, have become popular in development economics. However, RCTs do not always offer sufficient statistical power, which is reduced in experiments with imperfect compliance to treatment assignment. This is of critical importance if effect sizes are modest, and if non-compliance may occur. Both are likely in experiments in center-based childcare programs with individual-level randomization for several reasons. Dropout in the treatment group may occur because families' demand for preschool is unknown when the sample is constructed, and it evolves over time as households experience shocks and as they learn about the center. Non-compliance in the control group arises when children access the program or alternative preschool programs. This paper uses a recent evaluation of the Hogares Comunitarios program in Guatemala to illustrate challenges inherent in experimental evaluations and offers strategies to identify situations in which studies are more likely to succeed.

JEL classifications: C93

Keywords: Early childhood development, Research methodologies, Field experiments, Guatemala

* The author would like to recognize extensive and insightful feedback from Julián Cristia; this paper is vastly improved thanks to his comments. José Rodas, Jorge Matute, Olga Torres, Tania Santisteban, Julieta López and Carlos González generously shared valuable, honest information about the fieldwork. Horacio Alvarez provided insights on the program. Patricia Castañeda and Lilian Marroquín, as well as numerous staff members and *madres cuidadoras* from Hogares Comunitarios also generously shared their time and experiences, which made this paper stronger. Caridad Araujo, Jere Behrman, Santiago Cueto, Florencia López Bóo, José Rodas and Aimee Verdisco provided valuable feedback on the paper, which is appreciated. The Hogares Comunitarios impact evaluation discussed in this paper was made possible by funds from the Norwegian Trust Fund for Consulting Services.

1. Introduction

Randomized controlled trials (RCTs) have a long history in the biological sciences and other disciplines, and have become increasingly important in development economics (Duflo and Banerjee, 2011; Karlan and Appel, 2011). The experimental approach is appealing because of the robustness of the results of successful experimental evaluations; under simple and testable assumptions, RCTs generate unbiased estimates of treatment effects. Implementing an RCT is a risky endeavor, though, requiring significant upfront investments of time and money. Furthermore, researchers have few venues, and little incentive to share experiences of failed experiments. Because of this, there is little collective learning about when and how to run experiments effectively. Using a recent experimental evaluation of the Hogares Comunitarios childcare program in Guatemala as a case study, this paper analyzes challenges to implementing an RCT that are inherent in the evaluation of center-based care programs, and identifies under what circumstances an RCT is a safer bet. While this paper is focused on center-based childcare, many of the lessons learned and conclusions drawn here are relevant to experimental evaluations of other programs.

A useful impact evaluation offers estimates of treatment effects that are both unbiased and have sufficient statistical power to identify a treatment effect when one is present. RCTs are prized because they generate unbiased estimates when properly implemented under simple assumptions. What RCTs do not solve, however, is the problem of statistical power. With large sample sizes, large potential treatment effects and perfect compliance to treatment assignment, statistical power is not an issue. This is unlikely to be the case, however, in evaluations of center-based childcare programs. First, the effect sizes are likely to be moderate. Engle et al. (2011) find a median effect size of preschool programs of 0.24 standard deviations; with less structured educational components, childcare programs may have smaller effects. Detecting an effect size of 0.24 standard deviations might require a sample of 550 children (assuming individual-level randomization and an evenly divided sample). The analyst would be wise to use a study design that could detect a smaller size as well. Secondly, statistical power falls steeply as more participants fail to comply with their treatment assignment. Following the example of the sample of 550 children, if just one in four children defects from his treatment assignment, the minimum detectable effect size (MDE) doubles from 0.24 to 0.48 standard deviations, which means that the study would not permit the detection of effect sizes smaller than 0.48 standard deviations.

Maintaining an MDE of 0.24 with this level of non-compliance would require quadrupling the sample size to 2,200 children. Taking this into consideration, the analyst again would be wise to increase her sample size to accommodate the possibility of imperfect compliance.

Studies of center-based childcare are unlikely to escape the issue of imperfect compliance for several reasons that are described in detail in this paper. Imperfect compliance in the treatment group occurs when individuals have weak or unstable demand for the program. Even families that signed up for a program or are on a waiting list may not have strong and stable demand for preschool. Household shocks may reduce a family's demand for childcare: a parent may lose a job and prefer to stay home with the child, a family member may become available to care for the child at home, or the child may be needed at home for domestic work. Additionally, the childcare center may not meet the family's expectations. Center-based care is an "experience good;" like other experience goods, such as going to a movie or on a cruise, the consumer does not know the value of the good until one has begun consuming it. In the case of childcare, families do not know how well the child will adjust to the new environment and people at the center until the school year has already begun. The center may be a poor fit for the child if the child does not get along well with the teacher or the other children, or if unanticipated logistical difficulties arise that increase the cost of attending. When a child that drops out is in the treatment group of an RCT, this results in imperfect compliance. If a child from the control group takes the spot of the child from the treatment group, the compliance worsens further.

While compliance in the treatment group depends on the strength and stability of their demand for the program, compliance in the control group depends on excluding children who, by nature of the random assignment, have demand for the program that is just as strong as the demand of children in the treatment group. It may be politically and socially difficult to exclude children from the program, but not impossible with program staff that are willing to serve as effective gatekeepers. What is generally not feasible is to keep children from the control group from enrolling in an alternative childcare program. For this reason, the counterfactual is more likely to be outcomes for children in a combination of alternative childcare arrangements, rather than outcomes for children without access to any center-based childcare.

The Hogares Comunitarios study was discontinued because of non-compliance to the random assignment in both the treatment and control groups. As a consequence, the study lost precision, and the minimum detectable effect size of the estimated treatment effects rose from

0.17 standard deviations under perfect compliance to 1.32 standard deviations—an effect size that the program could not be expected to produce. Many children assigned to the treatment group failed to enroll in the program either because the family was never strongly interested in the program or because their situation changed. Others dropped out of the program over the course of the year because changes in the family’s employment or living situation meant that the family no longer needed childcare outside the home. Caretakers running the *Hogares* were gatekeepers that played a key role in determining compliance to the treatment assignment. They did not enforce the treatment assignment consistently for several reasons. First, many did not fully value the study or understand the importance of the random assignment. Second, in some cases, they felt obligated to enroll children from the control group out of concern for the children’s welfare. Third, caretakers’ pay was tied to full enrollment in their *hogar*, creating financial incentives that made it difficult for them to hold spots for children in the treatment group if children in the control group could begin attending sooner.

Several strategies might have improved the implementation of this intervention, and might be applied to others experimental evaluations. Communication problems might have been reduced if the research team presented the project in clear terms to caretakers, families and local program officials in a public forum. This might also have been beneficial in generating “buy-in” from key stakeholders in the study. Finally, financial incentives might have been modified to reward caretakers for enforcing the treatment assignment. Even with all these strategies, this might not have been enough to overcome the challenges associated with implementing a randomized controlled trial in a highly dynamic setting over the course of a year. RCTs are a high-risk, high-reward investment; if researchers can identify and minimize risks, they will be more likely to succeed.

The rest of this paper is organized as follows. Section 2 provides an overview of the role of experimental evaluations. Section 3 provides a more detailed discussion of the unique challenges associated with conducting a randomized evaluation of center-based childcare programs. Section 4 illustrates these issues using the Hogares Comunitarios study as a case study. Section 5 provides lessons from other experimental evaluations. Section 6 concludes.

2. The Role of Experiments

Randomized field experiments' increasingly prominent role in development economics is well known (Ehmke and Shogren, 2010; Bardsley et al., 2010; Smith, 2003). Economist Esther Duflo was awarded the prestigious Bates Clark Medal in 2010 in recognition of her work using field experiments to test economic theory in the field. Two prominent research institutes, the Jameel Poverty Action Lab at MIT and Innovations for Poverty Action at Yale have coordinated hundreds of experiments around the world. In a recent review of the literature on impact evaluations of interventions in education in developing countries from 1990-2010, Glewwe et al. (2011) found that more than half of published studies based on RCTs were published in the last three years alone.

For researchers interested in identifying policies' and programs' causal effects, the RCT is an important tool; under ideal circumstances, RCTs offer robust estimates of programs' causal effects on outcomes of interest without relying on elaborate econometrics or assumptions that are not directly testable. Furthermore, the results are intuitive for policymakers that may not have a background in the latest econometric methods. Policymakers increasingly rely on the results of experimental evaluations to inform policy. Mexico's *Progresa* conditional cash transfer program was one of the first large-scale evaluations of a public program. Following in Mexico's footsteps, numerous other countries implementing CCT programs have conducted experimental evaluations of the programs (Fiszbein and Schady, 2009).

The experimental approach has also been the subject of criticism (see Ehmke and Shogren, 2010; Barrett and Carter, 2010; and Card, DellaVigna and Malmendier, 2011 for recent discussions of these critiques). There are several reasons why RCTs are not a panacea for researchers interested in understanding programs' impacts. While RCTs' strength lies in the strong internal validity of their results, the external validity of their results is not assured. Furthermore, while RCTs generate robust estimates of a program's impact, the methodology does not shed light on *how* a program generates effects. Furthermore, implementing an RCT may be more costly in terms of time, human resources and financial resources than conducting a non-experimental retrospective study with existing data.

A practical drawback to RCTs is that experiments are generally high-cost, high-risk and may "fail" for a number of reasons after significant upfront investments of time and money have already been made. First, external events may interrupt the program being studied or data

collection. There is little a researcher can do to protect a study from unexpected political upheaval or natural disasters. On a local level, it may also be difficult to anticipate events such as the turnover of key staff people in the program being studied. Additionally, participants may react unexpectedly to the study or the program—simple interpersonal conflicts can negatively impact a study. Second, problems in implementation can derail a study. Assumptions about time or cost of implementation can be incorrect, as can researchers' expectations of how participants will react to the study. Fieldworkers can make mistakes in implementing a well-designed study. Third, as is discussed below, the study design must be sufficiently precise to detect the effects that the program is likely to generate.

As the experimental approach grows in popularity, research funds are increasingly dedicated to RCTs. Many field experiments are likely to yield important results that are surely worth the investment. Many others, however, will suffer setbacks significant enough that they will be abandoned before completion, often at great expense. If researchers are able to identify under what circumstances experiments are most likely to fail, and either employ strategies to minimize these risks or avoid investing in experiments that are high risk and high cost, scarce research funds may be allocated more efficiently.

3. Challenges Unique to Evaluating Center-based Childcare Programs

Despite the popularity of RCTs and the importance of early childhood development, there is very little experimental evidence of the effect of center-based childcare programs or preschool programs. In the United States, the Perry Preschool project and the Abecedarian Project are two highly influential randomized experiments (Schweinhart et al., 2005; Masse and Barnett, 2002). Conducted in the 1960s and 1970s respectively, these are the last widely cited experimental evaluations of preschool programs in which individual children were randomly assigned to attend preschool or not. Engle et al. (2011) conduct a systematic review of evaluations of ECD programs in low and middle-income countries published since 2007, and find 15 studies of center-based care that met their selection criteria. Of these, four used experimental methods. In two of the four, a small number of schools were selected to receive additional services, while in the other two, randomization was done at the individual level (Raine et al., 2003; Rolla et al., 2006). In these studies, students were randomly assigned to participate in one of multiple alternative programs; neither used a control group that received no intervention.

The dearth of experimental evidence on the effects of center-based care may be due to how difficult it is to implement an evaluation with sufficient statistical power to detect the effects that center-based programs are likely to produce. This is driven by two factors. First, effect sizes in educational interventions are not particularly large. Of the evaluations of center-based programs included in their review, Engle et al. find that the median effect size of attending preschool versus not attending is 0.24, while the median effect of making qualitative improvements to early childhood programs was 0.28 standard deviations. Evaluating programs with effects of this size requires a design with a minimum detectable effect size (MDE) that is small enough to detect these effects.

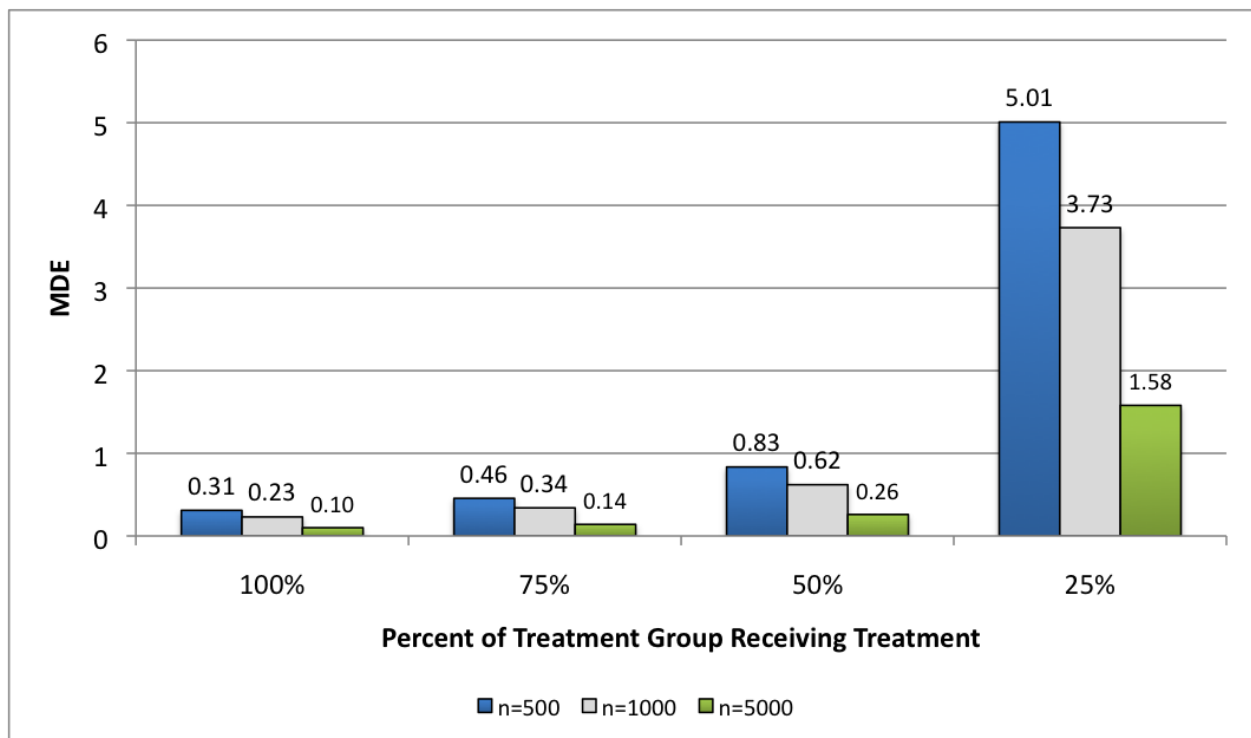
Two key determinants of a study's MDE are sample size and participants' compliance to their treatment assignment. The researcher seeks to maximize statistical power within his or her budget constraint. Data collection costs—likely to be the largest cost of a field experiment—generally increase with sample size. As such, the researcher will be interested in using the smallest sample possible while maintaining sufficient statistical power. In this situation, individual-level randomization is appealing because it results in greater statistical power than if the same number of individuals were randomized at the community level.

Maintaining the benefits of a large sample size, however, depends on participants' compliance to treatment assignment. Failure to comply with assignment to treatment in either the treatment or control group rapidly erodes a study's statistical power. Consider, for example, a study within which 500 children are randomly assigned to attend preschool or not. With *perfect* compliance in the treatment and control groups, the MDE for this study would be 0.25, similar to the median effect size of 0.24 found by Engle et al. for center-based preschool programs. With this design, if the program being studied has a similar effect size and *perfect* compliance, the researcher has roughly a 50 percent chance of identifying a significant effect with perfect compliance in the treatment group. Maintaining perfect compliance, and doubling the sample size to 1,000 participants, the MDE falls to 0.18, somewhat improving the researcher's chances of identifying an effect. Figure 1, below, gives examples of the MDE for samples of 500 or 1000 participants at various levels of compliance in the treatment group, assuming that 20 percent of the control group gains access to the program. In evaluations of center-based preschool programs, 20 percent contamination of the control group is not uncommon, given that even if

children are denied access to the program being studied, they are likely to be able to enroll in alternative preschool or day care programs.

Figure 1 demonstrates the rapid rise in the MDE associated with imperfect compliance. Even with perfect compliance in the treatment group and 20 percent contamination of the control group, the MDE is large enough that the researcher may not be able to detect an effect size similar to the median effect found by Engle et al. Under perfect compliance, the MDE for a sample of 500 children is approximately 0.25. Under perhaps more realistic assumptions, with 20 percent contamination of the control group, the MDE rises to 0.31. If one in four children in the treatment group drops out of the program, the MDE rises further to 0.46—nearly double the MDE under ideal circumstances. As can be seen in Figure 1, the effect of non-compliance dwarfs the effect of sample size.

Figure 1. Minimum Detectable Effect Size with 20 Percent Contamination of the Control Group and Varying Levels of Compliance in the Treatment Group



There are several reasons why perfect or nearly perfect compliance is difficult to achieve in an evaluation of center-based ECD programs with individual-level randomization. On a basic level, experiments involving human subjects differ from RCTs in the biological sciences because

researchers do not directly control their subjects' actions. Without being able to directly control the participants' actions, the researcher would like to identify a sample of individuals with homogeneous, strong demand for the program, and would like to have an effective exclusion mechanism that keeps individuals assigned to the control group from receiving the treatment. The researcher can then randomly assign individuals to the treatment or the control group and expect that all individuals that are offered the treatment will accept, while all individuals that are denied the treatment will be effectively excluded.

Compliance in the treatment group depends largely on those individuals' demand for the program; after all, it is up to the families whether or not the children remain enrolled and attend the preschool program every day. The challenge for the researcher is that the families' demand for the program is not easily observed. An application process or some other logistical hurdle to sign up for the study or the program may filter out those that have only a passing interest in the program. Requiring both parents to sign an application form may be important if one parent signs up, only to discover that the other parent does not agree. An application process may also prevent future dropout if it serves to educate parents on the program at the time they sign up. This has only limited utility, however, because at the moment they sign up for the program, new families do not know what their own demand for the program is. This is because preschool programs are a type of "experience good." Parents and children will not know if the child likes the program until the program has begun. It is understandable that if a child cries every day or does not get along with other children in the program, the parents would stop sending the child to the preschool. For issues that are discovered only after beginning the program, even the best filtering mechanism will not keep future dropouts from enrolling. If the analyst has an idea of what characteristics predict dropping out, he may use a simple survey at baseline to determine which families to include in the randomization, excluding those that appear to have weak or unstable demand.

Demand for preschool depends not only on the quality of the program and whether it's a good "fit" for the child, but on household characteristics that are likely to change over time. For example, if both parents are working outside the home, a family's demand for preschool is likely to be high because of their need for someone to care for the child while the parents work. If a parent loses a job, or if another family member becomes available to care for the child at home, this family may decide to remove the child from the program to care for the child at home. This

is particularly likely if there is a cost associated with attendance. Although these costs are likely to be covered by the study, costs may include tuition, other school fees or transportation as well as the time it takes to take the child to and from the preschool.

Program employees can play an important role in ensuring the continuity of participation among members of the treatment group. If teachers or administrative personnel know that it is important for children in the treatment group to have consistent attendance, they may give these children priority, and follow up with families in the treatment group if they stop attending. Evaluations that utilize individual-level randomization at a large number of schools necessarily involve a large number of decision-makers if each teacher or caretaker may decide which children to admit. Managing each decision-maker, particularly if the sample is geographically dispersed, may be challenging.

Whereas maintaining compliance in the treatment group depends largely on the strength and stability of families' demand for the program, maintaining compliance in the control group has more to do with program officials keeping individuals in the control group from accessing the program; this requires strong support from program officials and staff. They must understand that they are not to admit children in the control group, and ideally, they should understand why. If program officials do not understand and value the role of the study—or if they face incentives that are not consistent with denying children in the control group access to the program, there is no reason to expect that they would not prioritize the desires of the families in the control group over the requirements of the study. Winning over program officials and staff is a challenging task that is made easier in most circumstances by working with a small group. Working with a small number of relatively large centers rather than working with a large number of small centers is one strategy to minimize the number of staff and, therefore, the number of gatekeepers involved in the study.

If individuals in the control group are aware that they are being denied participation in preschool or some other benefit, it may be beneficial to offer them some alternative benefit, even if it is a simple token of appreciation. They may make individuals in the control group more willing to participate in data collection, and may make program officials in a gatekeeper role feel more comfortable with excluding them from the program.

In some situations, it is not ethical to exclude a child from treatment. For example, if a program has a strong feeding component, it is not ethical to exclude a child suffering from

malnutrition from the program to make room for a relatively healthy child that was assigned to the treatment group. Excluding malnourished children from the study sample and, therefore, from the randomization, is a strategy to avoid having to violate the treatment assignment for ethical reasons. This may not be straightforward in practice, however, since it may not be possible to identify malnourished children when constructing the sample. Furthermore, children that are healthy at the beginning of the study might become malnourished over the course of the study period.

While it may be possible to keep the control group from accessing the program being studied, it is less likely to be possible to exclude the control group from *all* preschool programs, particularly if the study is conducted in an urban area where alternative preschool or day care options are available. Individuals in the control group will be less likely to participate in an alternative program if their demand for preschool is highly price elastic (or highly sensitive to changes in price). If individuals have highly price elastic demand, this means that if the cost of attending preschool is lowered for individuals in the treatment group, these individuals will be much more likely to attend than individuals in the control group, who face some higher cost. Targeting low-income populations will increase subjects' elasticity of demand and reduce contamination in the control group.

Compliance to treatment assignment is complex, and relies on a combination of characteristics of the participants, the context, and program officials. Table 1 summarizes desirable characteristics of an RCT of center-based childcare with individual-level randomization, and identifies which are under the researcher's control. Table 2 summarizes the individuals that have a role in determining compliance to treatment assignment.

Table 1. Desirable Characteristics for an RCT of Center-based Childcare

| Desirable Characteristics | Under Researcher Control |
|---|---|
| Participants' demand is known | Limited control with a baseline survey |
| Participants' demand is stable | Limited control with a baseline survey |
| Participants' demand is highly price-elastic | Partly (by targeting low-income participants) |
| Effective mechanism to exclude the control group | Yes |
| Control group is unlikely to enroll in similar alternative programs | No—except by working in (probably rural) areas with no alternatives |
| Control group does not feel entitled to treatment | Partly (through targeting) |
| Participants may not opt out easily (e.g., by dropping out) | No |
| There are few gatekeepers; study team can influence their behavior | Somewhat—if feasible with small number of large centers |
| Benefits for control group | Yes |
| Static situation after initial treatment assignment | No |

Table 2. Levels of Compliance

| Level | Role |
|------------------------------|--|
| Administrative | May advise center-level employees to support the study by enforcing treatment assignment. |
| Teacher / Childcare provider | Decides whether or not to enforce treatment assignment by enrolling and following up with children in the treatment group, and by not enrolling children in the control group. |
| Parent | Decides whether or not to enroll the child and whether or not to continue sending the child. |
| Child | Responds to the childcare center environment with their behavior at the center as well as in telling their parents if they like going to the center or not. |

Community-level randomization eliminates a number of the challenges associated with individual-level randomization. Excluding control group children from the program is a simpler task since the program is not made available in their community. This facilitates the gatekeepers' job of excluding the control group. Potential discontent in the control group will be reduced, as they are less likely to feel entitled to have access to a program that is not available in their own community.

Imperfect compliance is likely to still be an issue, however. The treatment group's demand for preschool is still unknown and potentially unstable, so community-level

randomization does not eliminate the drop-out issue. Meanwhile, while members of the control group are unlikely to access the preschool being studied, in many settings, they will access alternative preschool programs.

A challenge associated with community-level randomization is that a larger sample size is required to achieve the same level of statistical power as would be attained with individual-level randomization. For example, the MDE of a randomized study with a sample of 1,000 children that has each been randomly assigned to participate in preschool or not is 0.18. The MDE increases to 0.30 if 50 groups of 20 children, totaling 1,000 children, are randomly assigned to attend preschool or not. Maintaining the MDE of 0.18 with clusters of 20 children requires a sample of 145 groups of children, or 2,900 children. Statistical power improves more by adding a large number of additional preschools with few children than by adding a few large preschools: an MDE of 0.18 can be attained with fewer than 2,900 children if the clusters are smaller. One hundred ninety schools with 10 children in each gives an MDE of 0.18. While it is advantageous in terms of power to work with a large number of small schools, this makes supervision and data collection more costly.

4. The Hogares Comunitarios Study

This paper explores some of the challenges associated with implementing experimental evaluations, with special emphasis on experiments in early childhood development programs in which the unit of randomization is the individual, using the recent experimental evaluation of the Hogares Comunitarios program in Guatemala as a case study. This experience offers lessons that may be useful for other researchers, particularly those considering similar evaluations.

This study draws on attendance data from the study, and interviews with the research team, field staff, Hogares Comunitarios authorities and staff, and 10 individual caretakers from two departments. This section illustrates some of the general points outlined above with the example of the Hogares Comunitarios study.

Hogares Comunitarios (HC) is a nation-wide, publicly funded program run by the First Lady's Secretariat for Social Work (SOSEP for its name in Spanish) in Guatemala to provide full time day care, nutrition and educational activities for children aged 6 months to 6 years. The program targets low-income families, with an emphasis on families with single mothers in urban and peri-urban areas. The program operates either in a private home that has been equipped for

child care with a capacity of 10 to 12 children, or a larger center with multiple caretakers that can receive more children. Local women, referred to as “caretaker mothers,” run the homes. The caretakers are paid a modest stipend to provide childcare, food and educational activities to the children in their home five days a week, for up to 12 hours a day. Licensed teachers visit most of the homes and centers at least weekly to provide additional educational activities.

In the last months of 2009, a group of researchers in the United States initiated a discussion with the director of the HC program. The program seemed appropriate for an experimental evaluation for several reasons. First, the research was policy relevant and would be a valuable contribution to the literature. Similar programs are found in numerous countries throughout Latin America, including Mexico, Colombia, Peru and Bolivia among others. Research on the effects of the Hogares Comunitarios program in Guatemala would be informative to policy-makers interested in similar programs throughout the region. Existing literature on similar programs was based on quasi-experimental evaluations that rely on the use of non-random matched comparison groups (Ruel et al., 2006), instrumental variable techniques (Attanasio, Di Maro and Vera-Hernández et al., 2010), or propensity scores (Cueto et al., 2009; Behrman, Cheng and Todd, 2004); an experimental evaluation of a similar program would be a valuable contribution. Second, it seemed feasible to implement an experimental evaluation. Program officials and others in Guatemala that were familiar with the program had remarked that the program was very popular and that in many communities, there was excess demand for the program. Where there is excess demand, it is easier to construct a control group since some interested individuals will be able to enter the program while others will not, even in the absence of the experiment. The national director for the program supported the idea of a randomized evaluation.

The research team and the program director reached an agreement to move forward with the study in November 2009. The research team would select a sample of approximately 100 homes or centers around the country, allowing for a sample of approximately 900 children. For every home or center, field workers would work with the caretaker mother to identify children in the community whose parents were interested in enrolling the children for the next school year, beginning in January 2010. Children that would be offered a spot in the *hogar* would be randomly selected from a list of interested children. Baseline data would be collected in the first months of 2010, and follow-up data would be collected a year later. The research team would

evaluate the program's impacts on child growth, measured by anthropometric indicators; child development, measured by children's scores on six child development scales; and on the children's biological mothers' employment outcomes. The research team hired a reputable local firm with ample experience conducting fieldwork to implement the study.

Field workers visited approximately 200 centers and homes in eight departments around the country in December. In each home or center, fieldworkers described the study to the caretaker, asked if the caretaker was interested in participating and how many spots were still available for the following school year. If she agreed to participate, she would be asked to provide lists of children interested in enrolling. The lists were to include three times as many children as there were spots available; this way, the children could be divided into a treatment group, a control group, and a "buffer group" of children that could be admitted to the program if additional spots opened during the year without contaminating the control group. To encourage participation in the study, the fieldworkers promised a series of payments to the caretaker mothers to compensate them for their participation that would be made at four points in time: upon agreeing to participate, upon turning in the list of children interested in HC, when baseline data were collected, and at an unspecified future date, totaling US\$50. The final sample included 100 homes. To be included, homes had to meet three requirements: each home needed to have at least two spots available that had not already been promised for the following school year, the caretaker had to agree to participate in the study, and the caretaker had to turn in the list of interested children.

Homes were selected for the study during the month of December, the last month of the school year. It was at this point that caretakers began creating their lists of children interested in HC. This was a difficult time to do this work in part because, as previously mentioned, by this time, some caretakers had already promised all the spots in their home for the following year. Of those that were participating, many compiled the lists of interested children when their end of the year vacation had already begun. Given this very tight timeline, the fieldwork team had to work quickly to make the initial contact with the homes before vacation started, before all the spots had been promised, and with time still for the caretakers to identify sufficient interested families to compile a list of children from which to conduct the random selection. On average, the fieldworkers visited two to three homes or centers per day. The homes were located in eight

departments around the country—after significant travel time, this left a limited amount of time to spend at each home.

Some caretakers found it difficult to recruit enough children to fill lists with three times as many children as they had spots, and lamented that they would find it difficult to then return to the majority of the families later to inform them that there was not a spot available for their child. To address this concern, the caretakers were told that those families that were not selected would “benefit” as well as the families that were selected. The benefit the families in the control group would receive was a basket of food worth US\$2.70. The financial incentive may have also been important in encouraging the caretakers to produce lists as long as those that were requested.

After the initial communication, most of the subsequent communication was done by phone. This was facilitated by the fact that all of the caretakers in the study had cell phones. This proved to be an advantage because it allowed the fieldworkers to communicate with all the caretakers much more quickly than if they had visited them all in person. This was beneficial especially when it was time to inform the caretakers which children had been selected for the study. Despite the convenience afforded by phone communication, it later became clear that some of the caretakers found the phone communication to be inadequate. In interviews, several caretakers commented that they found it impersonal, and blamed it for misunderstandings.

The final lists were compiled in early January. Children between the ages of 6 and 48 months were considered eligible for the study. Eligible children were placed in “priority lists” in random order. Children that appeared at the top of the list would be offered available spots before children at the bottom of the list. Siblings were always placed together on the lists, since families were likely to want their children to have the same childcare arrangement. Caretakers were asked to admit children in the order in which they appeared on the list. The first 347 children at the top of the lists were considered the treatment group, while the last 569 children were considered the control group, giving a total sample size of 916 children. Children that did not meet the study’s age requirements were included in the list but were not included in the study.

Hogares Comunitarios homes and centers opened for the 2010 school year in mid-January. In the first days of the school year, fieldworkers called caretakers to find out which children were enrolled and attending. They learned that an important number of children in both

groups were not compliant to their treatment assignment: 30 percent of children in the treatment group were not enrolled, and 16 percent of children in the control group were enrolled and attending. In interviews conducted with caretakers after the study had ended, caretakers offered several explanations for why children that had been assigned to the treatment group were not enrolled. In some cases, their parents' employment situation changed, and they were no longer interested in childcare outside the home. Some families moved. In other cases, families claimed that they had been interested in participating in a study, but not in enrolling in an Hogar Comunitario. A small number of caretakers claimed that they had not understood that the children on the list they turned in would be randomly selected to enter the HC; in these cases, caretakers disregarded the priority lists and made their own determination of whom to admit. In some cases, caretakers followed the lists, but enough children from the treatment group failed to come that children from the control group were admitted to fill the spots. In other cases, the caretakers chose to prioritize individual children in the control group that they perceived to be especially in need of the program due to malnutrition or extreme poverty. Caretakers only received their full salary from Hogares Comunitarios if they maintained full enrollment; because of this, they had an incentive to enroll the first children they could find that were interested in the program.

Fieldworkers remained in regular communication over the phone with caretakers through May of the school year, obtaining attendance information for children in the study. By this point, non-compliance in the treatment group worsened. Fifty-five percent of children in the treatment group were not enrolled, and 22 percent of children in the control group were attending. Communication with caretakers was less frequent between May and August.

In September of 2010, the research team observed that the non-compliance to the treatment assignments among the sample remained at unacceptably high levels and initiated a more aggressive strategy to attempt to re-enroll as many children in the treatment group as possible to "rescue" the study. The research team stopped working with the fieldwork firm, but hired one of the workers to continue gathering monthly attendance data from caretakers as well as directly from families. This work was primarily done over the phone, but he also made personal visits to caretakers and families he could not reach by phone. Two new workers that were knowledgeable about early childhood education and had experience running trainings with populations similar to the caretakers began overseeing the effort. A key component of the

strategy was to hold half-day workshops in each of the departments around the country for the caretakers. The objective of these workshops was to regain the caretakers' support for the study by explaining why the study would be valuable and how their participation was necessary for it to be a success, explain how it worked and what their role was, gather updated attendance information and to distribute lists of children in the treatment group that they should try to re-enroll in their home or center. At this point, the local program authorities became actively involved and became advocates and liaisons for the study at a local level.

During these workshops, some misunderstandings became apparent. Some families misunderstood how they would benefit from participating in the study. Families were offered modest compensation for participating in the baseline survey, but some expected ongoing benefits. This may have been a misunderstanding based on the surveyors' explanation that they would return a year later to collect data again.

These initial misunderstandings were not discovered until several months into the study when the workshops were held. Caretakers and families were not in regular contact with representatives from the study throughout the year to get their questions answered. Although they had the contact information to reach the fieldwork firm, neither the caretakers nor the families called the contact numbers for the study with questions. Local program authorities were not in contact with anyone from the fieldwork company or the research team to be able to answer caretakers' or families' questions. This is an example of a challenge resulting from having a large and geographically dispersed sample.

Table 3. Attendance by Treatment Group and Month: January 2010-January 2011

| | 2010 | | | | | | | | | | | | 2011 |
|-----------|------|------|-------|-------|-----|------|------|------|-------|------|------|------|------|
| | Jan. | Feb. | March | April | May | June | July | Aug. | Sept. | Oct. | Nov. | Dec. | Jan. |
| Treatment | 70% | 62% | 46% | 44% | 45% | 44% | 43% | 41% | 30% | 36% | 36% | N/A | 39% |
| Control | 16% | 18% | 28% | 21% | 22% | * | * | * | * | * | * | * | * |

*The study team stopped consistently collecting data on children in the control group after June.

After the workshops, most caretakers had a better understanding of what the purpose of the study was. Many even demonstrated enthusiasm for the study and were eager to help in any way they could while others remained upset about the benefits that never materialized. Local program authorities began playing a key role, as they were valuable liaisons between the

research team and the caretakers. Despite the energy that the caretakers had for the study, there was little they could do. Many homes had no open spots to offer to children in the treatment group that had stopped attending. Furthermore, most children in the treatment group that had dropped out had found another child care arrangement, or had dropped out of the program because they no longer needed childcare outside the home.

At the end of the 2010 school year, it was clear that increasing compliance was going to be difficult, but the energy that the program authorities and caretakers had for the study offered some promise. The research team gathered all the data available on the children in the treatment group and placed each child into one of several categories based on available data: enrolled; not enrolled, but interested (“recoverable”); and not enrolled and not interested (“not recoverable”). One of the women that ran the workshops and one new person were hired to work with the families and caretakers of the “recoverable” children to re-enroll them in HC for the 2011 school year. If enough could be re-enrolled, the study would be worthwhile.

Another challenge became apparent at this point. The caretakers’ salary depended on maintaining full enrollment. If their home or center did not operate at full capacity, their salary was reduced. Because of this, it was not feasible for the caretakers to reserve a spot for a child in the treatment group if another child was ready to begin right away. To address this, the research team offered to compensate the caretakers for any wages lost due to reserving spots for children in the treatment group. An additional strategy was to hire teachers (itinerant educational specialists that visit one or two homes or centers a week to provide additional educational services) to do home visits to treatment children’s families to encourage them to reenroll in the program. By the end of the month of January, some treatment children had been enrolled. During the same time period, the same number of children from the treatment group had dropped out. It became clear that the non-compliance was too severe to be able to identify a treatment effect. With perfect compliance to treatment assignments, this study would have been able to detect effects as small as 0.17 standard deviations. Using the same parameters, with approximately two thirds of the treatment group not attending, and an estimated 20 percent of the control group enrolled in the program, the study would only have been able to detect effect sizes as large as 1.32 standard deviations, which was unlikely to be observed.

Carrying out the evaluation would have presented a number of further complications. Estimates of the treatment effect on the treated based on comparing outcomes for children that

did attend HC with children that did not, regardless of their treatment assignment would have been biased. While the treatment and control groups were very similar on a number of observed characteristics at baseline, children from the treatment group that chose to drop out differed significantly from children in the control group that opted into the program despite being assigned to the control group. Table 4 illustrates several of these differences.

Table 4. Child and Family Characteristics by Attendance in Month Five

| Characteristic at baseline | Mean (Attending) | Mean (Not attending) | P-value of difference |
|-----------------------------------|-------------------------|-----------------------------|------------------------------|
| Child under 1-year | 19.8% | 32.4% | 0.0162 |
| Mother works | 38.0% | 26.4% | 0.0003 |
| Mother is <i>Ladina</i> | 29.2% | 36.6% | 0.0324 |

Those that chose to participate in HC are likely to have higher demand for the program; they are also likely to benefit more from the program than those that chose not to participate. An estimate generated by comparing these two groups of children would be likely to suffer from upward bias.

An alternative approach would be to follow the original treatment assignment regardless of children’s actual participation in the program, using an intent to treat estimate. This is a popular method to estimate a lower bound for a program’s effectiveness because it is based on a comparison of two groups that both include children that have been exposed to the program as well as others that have not. In the HC study, in May, 45 percent of children in the treatment group were attending, while 22 percent of children in the control group were attending.

Given the limitations of estimates that would be possible given the study’s weak power, it would have been difficult to justify the high financial and time cost of collecting follow-up data

5. Other Experimental Evaluations—When Have They Worked?

For reasons explained previously, characteristics of center-based childcare programs may make them difficult to evaluate with individual-level randomization, largely due to the likelihood of imperfect compliance to the treatment assignment in the treatment as well as control groups. Several of these problematic characteristics were present in the experimental evaluation of the Hogares Comunitarios program, and contributed to the imperfect compliance that eventually led to the discontinuation of the study because of low power. Here, we look at several other studies—some early childhood programs and others that are not—to further understand under

what circumstances experimental evaluations of center-based childcare programs are likely to be viable.

5.1. The Head Start Impact Study

The 2010 Impact Study of Head Start, the national federally-funded program designed to improve the school readiness of children living in poverty, is one recent example of a randomized evaluation of a center-based childcare program. In 1998—years after the results of a quasi-experimental evaluation of Head Start found that the program had little to no benefit for participating children—Congress mandated a randomized impact evaluation of Head Start. The study sample was nationally representative, including 5,000 newly entering children at 84 locations around the country. Like the Hogares Comunitarios study, the Head Start study restricted its sample to centers that had waiting lists in order to construct control groups. Randomly selected individual children were offered a spot at their local Head Start, while others were not offered a spot. This study, too, faced imperfect compliance. Seventy-seven percent of 4-year-olds and 84 percent of 3-year-olds in the treatment group attended Head Start. Meanwhile, 18 percent of 3-year-olds and 14 percent of 4-year-olds attended other Head Start locations, and 25 percent and 33 percent of 3 and 4-year-olds attended other center-based childcare. Despite the partial compliance, the researchers were able to identify some significant benefits of the Head Start program for participating children (U.S. Department of Health and Human Services, 2010). Had the compliance been more complete, additional program effects might have been detected.

5.2. Jamaica Home Visits

Grantham-McGregor et al. (1991) conducted an individual-level RCT of an ECD program in Jamaica in which community health aides visited the homes of stunted children aged 9 to 24 months every week for two years. Participants received nutritional supplementation, stimulation, or both. Both the nutritional supplementation and the stimulation had significant effects on children's cognitive development when offered individually, and the combined intervention had a greater impact than either one. Households in the treatment and control communities received regular visits from medical personnel. Like the Hogares Comunitarios study, this study used individual-level randomization and focused on disadvantaged urban populations. There are several key differences, however, between this home visit study and the Hogares Comunitarios

study that reduce the likelihood that of non-compliance in the home visit study. First, members of the control group did receive some benefits, which may have diminished potential resistance to the randomized design. Second, although the intervention lasted two years, it did not require ongoing proactive participation from members of the treatment group like participating in the Hogares Comunitarios program. If participating families experienced a change in employment or some other household shock (other than moving), they could continue participating in the treatment. Third, demand was relatively stable. This is in part because household demand for the program did not depend on household shocks, and partly because children were less likely to reject the program than they were to reject Hogares Comunitarios (receiving nutritional supplementation or even an hour-long visit is less of an experience good than a 60-hour per week day care program). Fourth, community health aides, in their role as “gatekeeper,” are likely to have faced less resistance. To effectively exclude children from the control group from home visits, community health aides may have faced less resistance than the Hogares Comunitarios caretakers, who were asked to turn away families that requested the service. Finally, demand for the home visits is likely to be less elastic than demand for Hogares Comunitarios or other center-based preschool programs. Demand for Hogares Comunitarios could be very high if neither parent is available to look after the child at home, or very low if someone is available. These factors appear to have contributed to a successful evaluation with nearly perfect compliance (two children were lost from the stimulated group) and a small sample (n=129).

5.3. One Laptop Per Child Evaluation in Peru

Researchers from the Inter-American Development Bank in Washington, DC, are implementing an RCT to evaluate the One Laptop Per Child program in Peru. In this intervention, the randomization was conducted at the school level, with 209 schools receiving laptops and 109 receiving no intervention. This study has enjoyed almost perfect compliance. Several characteristics of this study’s design may explain this. First, the treatment was assigned at a school level, and implemented by the central Ministry of Education authorities, improving likelihood of compliance to the treatment. Elasticity of demand for the laptops is high, given that schools that are offered the laptops were likely to accept them, and schools that were not, are very unlikely to buy them or otherwise gain access to them on their own. Second, schools’ demand for laptops is likely to be more stable than families’ demand for preschool. Even if

teachers do not appreciate the laptops' value, schools are not likely to suffer shocks that will make them reject laptops. Furthermore, teachers' supervisors may ask them to use the laptops, further reducing the likelihood of non-compliance in the treatment group.

Table 5, below, reviews how Hogares Comunitarios and the three other RCTs described here compare against the desirable characteristics of an RCT that were identified in Table 1.

Table 5. Desirable Characteristics for an RCT: Examples

| | Hogares Comunitarios | Head Start | Jamaica Home Visits | One Laptop Per Child |
|--|-----------------------------|-------------------|-----------------------------|-----------------------------|
| Known demand | No | No | No | No |
| Stability of demand | Low | Moderate | High | High |
| Elasticity of demand | Low | Low | High | High |
| Exclusion of control group from specific program | Possible, though difficult | Impossible | Possible | Possible |
| Exclusion of control group from alternative programs | Impossible | Impossible | Possible | Possible |
| Control group does not feel entitled to treatment | False | False | True | True |
| Participants may not opt out easily (e.g., by dropping out) | False | False | False, though not a problem | True |
| Few decision-makers | False (100) | False (84) | True (1) | True (1) |
| Benefits for control group | None perceived | None | Medical care | None |
| Static situation after initial treatment assignment | No | No | No | Yes |

6. Conclusion

As previously described, RCTs generate unbiased estimates of treatment effects, but do not always generate estimates with strong statistical power. When evaluating programs that are likely to have modest effect sizes, a study design with sufficient statistical power is essential. Compliance to the treatment assignment is a key component to the eventual statistical power that a study has. Unlike in laboratory experiments, field experiments are likely to suffer from imperfect compliance. Individual-level randomized experiments of center-based early childhood increase the likelihood of imperfect compliance.

- **Individuals' demand for preschool is unknown.** Maintaining compliance to the treatment assignment requires minimizing drop-out. Participants assigned to the treatment group are more likely to drop out if their demand for

preschool is weak. Filtering mechanisms such as short surveys to identify people with no alternative child care arrangement, for example, may help to identify individuals with high demand. Because preschool is an experience good, with a value that is unknown to the families until it is consumed, a filter will generally not identify families that only discover that they have low demand for preschool after the school year has already begun. In the Hogares Comunitarios evaluation, children dropped out if they did not get along well with other children or with the caretaker.

- **Demand may be unstable.** Families' demand may change when the household experiences any number of shocks. A family member may become available to provide childcare at home (a parent may lose a job, or a grandparent may move in).
- **Complete exclusion of the control group from the program may not be possible.** Excluding the control group from the specific program being evaluated is challenging when there are many individual decision makers, and when the control group is in the same community as the treatment group, as was the case with the Hogares Comunitarios study. An unwieldy 100 caretakers were enforcing the treatment assignment. By comparison, in the One Laptop Per Child study, one central office determined which schools would receive laptops, and children in the control group attended schools in other communities.
- **Exclusion from alternative programs is impossible.** Except in very few circumstances, it is impossible to prevent children assigned to the treatment group from enrolling in alternative center-based care programs. This issue decreased compliance to treatment assignment in the Head Start Impact study and is likely to be an issue in any experimental preschool evaluation.
- **Managing decision-makers is key.** With individual-level randomization, enforcing the treatment assignment depends on program staff's cooperation. In the Hogares Comunitarios study, caretakers cooperated by generating lists of interested children in response to a financial incentive at the beginning of the study. Later, they faced a financial incentive to fill all open spots in their

home since their salaries depended on full enrollment. When a child from the control group wanted to fill a vacant spot, this conflicted with enforcing treatment status. This type of situation may be managed by aligning staff's financial incentives with the objectives of the evaluation, and by having clear communication.

- **The distribution of the sample matters.** Maintaining clear communication with all decision-makers is easier if there are fewer decision-makers, and if they are accessible geographically. Monitoring a large, geographically disperse sample is costly in terms of time and money. This is even more challenging if communication by phone is not effective. Even so, clear, frequent communication alone is not likely to be sufficient if decision-makers face conflicting incentives.

A successful study also depends on implementation. This is important to maintain compliance, and to maintain good will when it is time for data collection, which often relies on the support of local program staff. The Hogares Comunitarios experience offers a few lessons.

Simple misunderstandings between field staff, program staff and families contributed to confusion and false expectations. When families were told that they would benefit from participating in the study regardless of whether or not their child was selected to participate in Hogares Comunitarios, many interpreted the word “benefit” to mean an ongoing stream of goods or services. Because field staff was not present in the communities after the baseline survey was implemented, this confusion was not clarified. If field staff were able to have more of a presence in participating communities, or if program supervisors had been involved earlier on to clarify misunderstandings, this issue might have been resolved sooner, and caretakers might have had a more positive attitude about the study.

While some misunderstandings are inevitable with a large-scale study and a diverse population, several strategies may minimize the likelihood of misunderstandings. Caretakers suggested holding community meetings in which the benefits associated with participating in the study were explained publicly; this would ensure that everyone heard the same message.

Could the Hogares Comunitarios study have worked? Several things may have been improved, though they might not have been enough. First, it would have been feasible to provide closer supervision of preschools and maintain more constant contact with a smaller number of

preschools in a more geographically compact area. A small number of large centers in the Guatemala City area might have been a more manageable sample, although the disadvantage of working in an urban area is that the control group has ready access to alternative programs. Second, transparent public meetings with all participating families might have prevented some misunderstandings and reduced discontent amongst participants that expected “benefits” that never materialized. At the same time, these may have backfired if families objected to the idea of random assignment. Third, a better filtering mechanism could have been used to identify families with stronger demand. Nonetheless, this would not have filtered out families who would drop out because a parent lost a job, or families with children that would drop out because of conflicts with other children. Fourth, if program authorities agreed, modifying caretakers’ pay structure such that they were rewarded for enforcing the treatment assignment and not simply for maintaining full enrollment, might have improved compliance. Financial incentives worked well when caretakers were asked to produce lists of children, and they may have been an effective strategy to maintain compliance as well.

Facing these challenges, is it possible to conduct a successful experimental evaluation of a center-based early childhood program? This relies on several key elements falling into place. An RCT is more likely to work if study participants’ demand for the program is strong and stable. The gatekeeper should be willing to enforce the treatment assignment; this is more likely if the gatekeeper faces financial incentives that reward enforcing the treatment assignment. If participants’ demand is highly elastic, subsidizing their participation for the treatment group or imposing a cost for the control group will help improve compliance. In studies located in rural areas, members of the control group are less likely to have access to alternative programs. The control group’s participation in the program being studied (as opposed to similar alternative programs) may be reduced if randomization is done at the community level, although this reduces power for a given sample size, and does not prevent the control group from accessing alternative programs.

Experimental studies of center-based preschool programs in the United States have demonstrated that interventions in early childhood can have large effects on children’s cognitive development, educational attainment, earnings and participation in crime that are very cost effective (Schweinhart et al., 2005; Masse and Barnett, 2002). Quasi-experimental evidence from developing countries suggests that center-based programs are likely to have large effects in low

resource settings as well (Behrman, Cheng and Todd, 2004; Glewwe et al., 2011). Successful experimental evaluations of center-based programs in developing countries would be valuable contributions to the ECD and human capital literature. A successful study will not only generate unbiased estimates of the effects of center-based care, but will have sufficient statistical power to detect the effects that are likely to be found. This is a challenging task that is more likely when a strong study design is in place.

References

- Attanasio, O., V. Di Maro and M. Vera-Hernández. 2010. "Community Nurseries and the Nutritional Status of Poor Children: Evidence from Colombia." London, United Kingdom: University College London. Manuscript.
- Bardsley, N. et al. 2010. *Experimental Economics: Rethinking the Rules*. Princeton and Oxford: Princeton University Press.
- Barrett, C., and M. Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections." *Applied Economics Perspectives and Policy* 32(4): 515-548.
- Behrman, J., Y. Cheng and P. Todd. 2004. "Evaluating Preschool Programs when Length of Exposure to the Program Varies: a Non-Parametric Approach." *Review of Economics and Statistics* 86(1): 108-132.
- Card, D., S. DellaVigna and U. Malmendier. 2011. "The Role of Theory in Field Experiments." NBER Working Paper 17047. Cambridge, United States: National Bureau of Economic Research.
- Cristia, J. et al. 2011. "The Short-Term Impacts of the One Laptop per Child Program in Peru: A Randomized Evaluation." Washington, DC, United States: Inter-American Development Bank. Mimeographed document.
- Cueto, S. et al. 2009. "Promoting Early Childhood Development through a Public Programme: Wawa Wasi in Peru. Young Lives Working Paper 51. Oxford, United Kingdom: University of Oxford, Department of International Development.
- Duflo, E., and A. Banerjee. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York, United States: Public Affairs.
- Ehmke, M., and J. Shogren. 2010. "The Experimental Mindset within Development Economics: Proper Use and Handling Are Everything." *Applied Economics Perspectives and Policy* 32(4): 549-563.
- Engle, P. et al. 2011. "Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries." *Lancet* 378(9799): 1339-1353.
- Fiszbein, A., and N. Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, DC, United States: World Bank.

- Glewwe, P. et al. 2011. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010." NBER Working Paper 17554.
- Grantham-McGregor, S. M., C.A. Powell, S. P. Walker and J.H. Himes. 1991. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *The Lancet* 338: 1-5.
- Karlan, D., and J. Appel. 2011. *More than Good Intentions: How a New Economics is Helping to Solve Global Poverty*. New York, United States: Dutton.
- Masse, L.N., and W.S. Barnett. 2002. A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention. New Brunswick, NJ: National Institute for Early Education Research. Available at: <http://nieer.org/resources/research/AbecedarianStudy.pdf>
- Raine, A. "Effects of Environmental Enrichment at Ages 3-5 years on Schizotypal Personality and Antisocial Behavior at Ages 17 and 23 Years." *American Journal of Psychiatry* 160: 1627-35.
- Rolla, A. et al. 2006. "Evaluating the Impact of Different Early Literacy Interventions on Low-Income Costa Rica Kindergarteners." *International Journal of Education Research* 45: 188-201.
- Ruel, M. et al. 2006. "The Guatemala Community Day Care Program: An Example of Effective Urban Planning." Research Report 144. Washington, DC, United States: International Food Policy Research Institute.
- Schweinhart, L. et al. 2005. "The High/Scope Perry Preschool Study through Age 40: Summary, Conclusions and Frequently Asked Questions." Ypsilanti, United States: High/Scope Press.
- Smith, V. 2003. "Constructivist and Ecological Rationality in Economics." *American Economic Review* 93(3): 465-508.
- United States Department of Health and Human Services. 2010. *Head Start Impact Study*. Final Report. Washington, DC, United States: United States Department of Health and Human Services.

Appendix

Appendix 1. Participating Children by Department

