DESIGNING EXPERIMENTS TO MEASURE SPILLOVER AND THRESHOLD EFFECTS

Sarah Baird
George Washington University

Aislinn Bohren
University of California, San Diego

Craig McIntosh
University of California, San Diego

Berk Özler
DECRG, World Bank

December 2, 2011

Abstract

This paper develops the technique of experiments designed specifically to study the nature of spillover effects between subjects. By first randomizing the intensity of treatment within each cluster and then randomizing individual treatment conditional on this cluster-level intensity, a novel set of research questions can be addressed. Not only do we gain direct evidence as the impact of the treatment on untreated units, but the experimental variation in the treatment intensity allows the researcher a straightforward way to observe saturation and threshold effects among treated and untreated units alike. We present a framework in which to back out a very rich set of parameters in the analysis of such an experiment, and examine the power implications of the randomized saturation design relative to more standard designs. We demonstrate that the randomization of saturations at the village level brings empirical benefits even when we are interested in spillovers in dimensions other than the one in the saturations were directly randomized. The technique is implemented using a Cash Transfer program in Malawi; we find evidence of beneficial spillovers that improve cognitive performance among untreated girls, and expenditures on girls prove very sensitive to the saturation of treatment at the village level. Social and household networks appear to be particularly strong conduits for beneficial spillover effects that improve school enrollment and decrease the risk of HIV/AIDS.

Keywords: Spillover Effects, Experimental Design, Cash Transfers, HIV/AIDS.
JEL Codes: C93, O22, I25

Corresponding Author: McIntosh, 9500 Gilman Drive, La Jolla CA, 92093-0519. ctmcintosh@ucsd.edu.

1 Introduction

For many years the theory and practice of randomized controlled trials have maintained the Stable Unit Treatment Value Assumption (SUTVA), under which outcomes may respond to an individual's own treatment status but not to the treatment of others. More recently, particularly as experimentation has expanded in areas such as public health where spillover effects are clearly predicted by theory, it has become increasing common to try to relax this assumption. Researchers have sought to uncover network effects using experimental variation across treatment clusters (Miguel and Kremer 2004, Bobba and Gignoux 2010), using within-cluster strata that remained untreated (Angelucci and De Giorgi 2009, Bobonis and Finan 2009, Barrera-Osoria et al. 2008, Lalive and Cattaneo 2009, Alix-Garcia et al. 2011), or exploiting plausibly exogenous variation in within-network treatments (Duflo and Saez 2002, Munshi 2003, Conley and Udry 2010, Beaman 2010, Babcock and Hartman 2010). While the use of within-cluster controls or incidental randomness within networks has become more common, this paper is part of a newer literature in which randomized experiments are designed explicitly to uncover specific forms of spillovers (see Duflo and Saez 2003, Oster and Thornton 2009, Gine and Mansuri 2011, and particularly Sinclair et al. 2011 who implement a similar experiment). We develop a framework in which to understand the costs and benefits of implementing such designs, and then present data from an experiment conducted on a cash transfer program in Malawi.[1]

In the experiment studied here, the fraction of eligible schoolgirls offered a cash transfer program was directly randomized at the cluster (village) level, leading to what we term a randomized saturation design. This structure provides an additional twist on the 'partial population' experiment suggested by Moffit (2000) and allows for a straightforward comparison of treated and untreated outcomes across the directly randomized distribution of treatment saturations. We present a simple theoretical environment in which SUTVA is maintained across but not within clusters, and consider the problem of estimating treatment effects across the distribution of the treatment saturation.[2] Working with no additional assumptions we can identify the average Intention to Treat Effect of the

[1] A formal development of designs in which the saturations of treatment are randomized is particularly relevant as many such studies are currently in the field, including work on budgeting spillovers across Mexican municipalities, and financial spillovers within informal social networks in Sri Lanka.

[2] For an alternative approach, see Rosenbaum (2007) who provides nonparametric estimators based on randomization inference that allow for the construction of confidence intervals without maintaining SUTVA across clusters, as we do.

program across the observed saturation distribution, as well as estimating how outcomes of agents offered and not offered the treatment alter with the cluster-level treatment saturation. The design therefore provides an unusually clear window onto the type of threshold effects motivated by Brock and Durlauf (2001). We can further invoke an assumption over the homogeneity of spillovers for all units that do not receive the treatment and combine this with an assumption about the function form of the saturation-specific effects to fully identify all of the parameters of interest, including the Treatment on the Treated exclusive of any spillover effects and the variation in the Treatment on the Treated induced by changes in treatment saturation.

Our empirical exercise uses the cash transfer experiment to examine three sets of outcomes: academic (enrollment and human capital), demographic (marriage, pregnancy, and HIV prevalence), and economic (expenditures on goods specific to adolescent girls). We present results pooling the various treatments together, as well as results that show the separate spillover effects for the two sub-arms of the transfer experiment: transfer conditional on school attendance and transfers given unconditionally (Baird et al. 2011). The empirical analysis begins with an exercise based rigorously on the dimensions directly randomized in the research design. We then move to a more general consideration of network effects, and demonstrate that even if the randomized saturation is conducted in the 'wrong' network, as long as there is any correlation between the chosen and actual network the additional dimension of randomization will still increase the variation in observed saturations relative to other designs. A Monte Carlo exercise illustrates this point by simulating the saturations that would have been observed in the social networks of our study subjects had we used a blocked, clustered, or randomized saturation method to assign treatment. We then exploit the experimental variation in treatment saturation within social networks and within households to look for evidence of spillover effects within these more tightly defined social groups.

Several literatures suggest that schooling cash transfer programs should be expected to have meaningful spillover effects. First, schooling decisions are likely to be influenced by the behavior of peers, whether because the value for education is a social norm (Akerlof and Kranton 2002) or because local investments in education are a club good (Benabou 1993). Conversely, school resources may be zero-sum, in which case increases in enrollment by treated students can lead to congestion effects that harm the outcomes of those who do not receive the program (Behrman, Sengupta, and Todd 2005). The cash transfers themselves can be an additional source of spillover effects; Angelucci

et al. (2010) find that intrahousehold transfers modulate the impact of cash transfer programs by allowing extended families to push resources towards secondary-school children who might otherwise have dropped out. Finally, we are interested in understanding spillover effects of the program on marriage, pregnancy, and HIV, all of which are directly subject to network effects (Obbo 1993, Behrman et al. 2003).

The empirical evidence on the spillover effects of Cash Transfer programs has been mixed. Among several papers investigating impacts on ineligibles from the Progresa experiment, Behrman et al. (2005) find no evidence of spillovers on the probability of enrollment, but Angelucci and De Giorgi (2007) find substantial impacts on consumption. Bobba and Gignoux (2010) instead relax the assumption of SUTVA across villages, and show that a higher saturation of treatment villages in the local area leads to improvements of enrollment for treated but not control units. Barrera-Osorio et al. (2008) examine a different cash transfer program in Colombia, looking for evidence of spillover effects within household and social networks. They uncover some evidence of negative spillovers for children who were enrolled but not selected in the lottery that determined treatment status when they had siblings who were selected. Within social networks, on the other hand, they find robust evidence of positive spillover effects. Macours and Vakis (2008) find evidence that the impacts on beneficiaries of a Nicaraguan cash transfer program vary according to proximity to treated elites, but Macours, Schady, and Vakis (2008) test for spillovers effects on ineligible households in Nicaragua and find none. Thus the empirical literature provides a series of snapshots of spillovers on different populations using different outcomes, and comes to no clear conclusion.

We designed our experiment to provide straightforward evidence on the spillover effects of cash transfer programs on beneficiaries and non-beneficiaries alike, over a broad set of outcomes. We gave Conditional Cash Transfers (CCTs) to some villages and Unconditional Cash Transfers (UCTs) to others so as to be able to distinguish the effect to which non-beneficiaries were effected by shifts in the emphasis on schooling rather than simply through income sharing. We track outcomes using a set of objectively collected data on school attendance (as reported by the schools themselves) and sexually transmitted diseases (biomarker testing of HIV/AIDS and Herpes Simplex Virus-2). The CCT program has been found to increase schooling, the UCT program to decrease marriage and pregnancy (Baird et al 2011a), and both programs have been shown to lead to a significant decrease in STDs (Baird et al 2011b) . Particularly because the improvements in outcomes such as HIV

could easily arise as a result of the diversion of sexual activity, careful research design on spillovers is critical. A program that caused significant decreases in HIV infection for treated individuals at the cost of a symmetric increase in untreated individuals would have no overall effect at all.

Our results indicate that the spillover effects of this Malawian transfer program were quite muted on average. Using the cluster (village)-level experiment, there are no significant differences between the within-village controls and the pure controls for any of our core outcomes. Despite this lack of average effects, expenditures on girl-specific goods display strong heterogeneity over the village-level treatment of saturation: the higher the fraction of girls treated, the more money goes towards untreated girls and the less is focused on the treated girls who become less 'special' when treatment is near-universal. There is weak evidence that the improvements in enrollment seen in the treatment group are smaller as the fraction of treated goes up, consistent with the presence of a congestion effect in schooling. Once we split conditional from unconditional villages, we see that the improvement in cognitive test scores among the within-village CCT control is as large as the treatment arm (and this effect is increasing in the treatment saturation), suggesting that more learning may have occurred in conditional villages even for those not directly treated. When we move to the analysis of spillovers within social networks, we find that untreated girls attend school .11 additional terms for each additional friend that is treated, and that the marginal effect of having conditional friends is twice the effect of having unconditional friends. Treatment within one's social network and household network appears to have a powerful effect on protecting girls from HIV, with one additional friend decreasing prevalence by 2% and one additional sister by 4%.

The remainder of the paper is organized as follows. Section 2 lays out the problems that arise in trying to estimate spillover and saturation effects using standard research designs, discusses the assumptions required to use a randomized saturation design to fully identify all the parameters of interest, and provides a closed-form expression for the decrease in power that arises from a randomization of treatment saturation. Section 3 presents the sampling strategy and data used in the paper and lays out the details of the research design. Section 4 examines spillover and saturation effects using the specifications directly motivated by the research design, while Section 5 presents estimates of spillover effects within household and social networks, as well as impacts of the program on the composition of the networks themselves. Section 6 concludes.

2 Theoretical Motivation

2.1 Sample Size and Spillovers using Standard Designs

We motivate the role of spillovers in research design by pointing out a conundrum that confronts the applied researcher. One of the most fundamental design choices in any multi-level experiment is the question of how to allocate treatment to N individuals distributed across J clusters. The bracketing research designs in this case are the 'blocked' design, in which exactly half of each cluster is treated, and the 'clustered' design, in which half of the clusters are assigned to full treatment and half to control. While it is typically logistically more straightforward to treat at the cluster than the individual level, the 'design effect' is the loss in power that arises from doing so. The estimation of intra-cluster correlations (ICCs) from baseline data is considered key to understanding the tradeoff inherent in this design choice, and the power advantages of a blocked design will be increasing in the component of correlation in outcomes that is within-cluster (see Duflo et al. 2007 for a concise summary of this issue).

Given that the presence of a high intra-cluster correlations might otherwise predispose us to pursue a blocked design, it is critical that we step back and ask ourselves why units within a cluster are behaving similarly. To use the terminology of Manski (1993), clustering may arise because of correlated effects (similar units congregate), contextual effects (behavior is driven by exogenous characteristics of others in the cluster), or endogenous effects (behavior is driven by the behavior of others in the cluster). This distinction is critical for research design, because endogenous effects will violate SUTVA while correlated and contextual effects will not. The 'reflection problem' refers to the difficulty of distinguishing these effects with the types of observational data to which the experimenter typically has access at baseline.

The presence of a high intra-cluster correlation, if generated by correlated or contextual effects, would incline us to the use of a blocked research design. If instead this within-cluster correlation arises due to an endogenous effect, however, then we show in what follows that the blocked design is not only biased but provides no information on the extent to which endogenous effects are to blame. Thus the applied researcher ends up on the horns of the reflection problem, and basic research design choices depend on the extent of spillover effects. In this sense the 'randomized saturation' design proposed here is closely related to the empirical estimation of network effects, since only if we have

an instrument to resolve the reflection problem can we uncover the underlying reasons why agents within a cluster appear similar.

While the clustered Intention to Treat effect (ITT) is not biased by spillover effects, it also does not permit us to measure them directly, nor can we back out the Treatment on the Treated (ToT) in the presence of spillovers with an instrumentation strategy as we can do when we maintain SUTVA. While we can conceptually decompose the ITT into a ToT experienced by compliers and an Average Spillover Effect (ASE) experienced by non-compliers, if there are any unobservable determinants of compliance then these quantities cannot be measured. This point is well understood in the literature, and consequently researchers who seek to measure spillovers directly have typically exploited what Moffit (2000) refers to as a 'partial-population experiment', whereby a subset of units within treated clusters are not offered the treatment.

By far the best-known example of such an experiment is Progresa/Oportunidades in Mexico, which has been used to study spillover effects driven through transfers within household networks (Angelucci and di Giorgi), peer effects in school enrollment (Bobonis and Finan 2009), and market-mediated production spillovers (Alix-Garcia et al., 2011). Progresa is a partial-population experiment because it features both a treatment selection decision at the cluster (village) level as well as an objective poverty eligibility threshold at the household level. Other recent examples from the literature include Duflo and Saez (2003) who examine enrollment in retirement plans within academic departments, and Kapteyn et al. (2010) who examine spillover effects of lottery winnings within Dutch postal codes. Such partial-population experiments are useful because the entire group offered treatment can be compared to the relevant pure control for an estimate of the ITT, and the within-cluster controls can be compared to the relevant pure controls for an estimate of the ASE.

Most extant partial-population experiments do not permit the estimation of two important relationships, however. First, when the fraction of the sample offered the program in treatment clusters is either fixed or endogenous, we have no direct way to estimate saturation or threshold effects, meaning the impact of varying the fraction of units offered treatment within treatment clusters. This quantity is of critical policy interest, because in contexts where strong cluster-level spillover effects are present, it may be completely ineffective to treat below some saturation level (see Barham and Maluccio 2009 for evidence on vaccinations, and Tarozzi et al. 2011 for evidence related to anti-malaria bed nets) or cost-ineffective to treat above some threshold. Virtually all

extant partial-population experiments feature cluster-level saturations that are either endogenous (such as in Progresa, where they are determined by village-level poverty) or fixed (such as in Duflo and Saez (2003), where they are set at 50%). Seen from this perspective, all spillovers estimated from such experiments can be thought of as a kind of Local Average Spillover Effect estimated at a specific saturation. If the actual implementation of the project were conducted with a different saturation level, then both the ITT and the ASE estimated by these limited experiments may be incorrect. Further, partial-population studies conducted with a single treatment saturation do not provide any information to policymakers as to the optimal saturation at which treatment should be conducted in practice.

In addition, many well-known partial population experiments exploit within-cluster controls that were not treated because they were ineligible for the program. Comparison of ineligibles in treatment versus control clusters does indeed permit identification of spillover effects under very weak assumptions, but there is no reason to think that the spillover experienced by ineligible units would be the same as the spillovers experienced by those who are eligible. The reason this distinction is important is that if we have a measure of average spillover effects within the population of eligibles we can back out an estimate of the Treatment Effect on the Treated even in the face of violations of SUTVA. In the next section we formalize the advantages of a Randomized Saturation design over the standard partial-population experiment and the assumptions required to gain an estimate of all the parameters of interest.

2.2 The Randomized Saturation Design.

Manski (2010) works in a completely general framework that permits arbitrary forms of spillovers between units, but in this general setting the domain of the treatment response becomes the Cartesian product of the treatment across all the individuals in the population. Causal inference in this context is challenging both because we cannot clearly identify 'control' units whose outcomes have been undisturbed by the treatment, and also because we cannot know the extent to which a unit has been treated without a complete mapping of the relevant network for a specific individual.

We now present a framework that is intended to capture the essence of spillover effects in many applied contexts, where research subjects are gathered into clusters that may feature extensive spillover effects within cluster, but in which spillovers across clusters are likely to be limited. This

allows us to maintain SUTVA across clusters while relaxing it within them. To simplify the problem further, we then focus our analysis of spillovers on the key parameter under the control of the policymaker, which is the fraction of eligible individuals within each cluster that will receive the treatment, or the 'saturation'. The value of this approach is that it permits us to design an experiment that reveals perhaps the two most important implications of non-independence for applied policy contexts: spillover effects, whereby the treatment status of one individual effects outcomes for another, and threshold effects, whereby cluster-level outcomes may be some discontinuous function of cluster-level treatment intensity.

To fix ideas, we augment the standard notation of the Rubin Causal Model (Rubin 1974) by writing outcomes as $Y_{ij}(T_{ij}, C_{ij}, S_j; X_{ij})$, where Y_{ij} is some outcome for individual i in cluster j, $T_{ij} \in 0, 1$ is an indicator for being offered the treatment and $C_{ij} \in 0, 1$ indicates compliance with the treatment, $S_j \in [0, 1]$ represents the cluster-level saturation of treatment among eligibles, and X_{ij} is a vector of covariates. In this framework, the Intention to Treat effect (ITT) is dependent on the cluster-level saturation because an RCT in which every eligible unit is treated may result in a different treatment effect than one in which only a single unit in each cluster is treated:

$$ITT(p) = E(Y \mid T = 1, S = p) - E(Y \mid T = 0, S = 0)$$

The non-independence between units implies that individuals in treated clusters may be subject to some spillover effect. In full generality we can think of the Average Spillover Effect as depending on the treatment and compliance status of an individual, as well as on the cluster-level saturations. We simplify the problem by assuming that spillovers (conditional on saturations) are the same for all individuals who do not receive the treatment (all non-compliers plus compliers in the control), and hence we have two quantities of interest; the Average Spillover on the Treated and the Average Spillover on the Non-Treated: Thus we can write

$$AST(p) = E(Y \mid T = 1, C = 1, S = p) - E(Y \mid T = 1, C = 1, S = 0)$$

$$ASNT(p) = E(Y \mid T * C = 0, S = p) - E(Y \mid T * C = 0, S = 0)$$

These quantities give the difference, conditional on actual receipt of the treatment, between

outcomes at saturation p and outcomes if no-one else had been treated. Having defined spillovers in this way, it is then natural to define a quantity we call the 'Treatment on the Uniquely Treated' as

$$TUT = E(Y \mid T = 1, C = 1, S = 0) - E(Y \mid T = 0, C = 1, S = 0).$$

In other words, the TUT is the Treatment on the Treated among compliers if only a single person in each cluster is treated. This allows us to decompose the saturation-dependent Treatment on the Treated into two additively separable terms:

$$\begin{aligned} ToT(p) &= E(Y \mid T = 1, C = 1, S = p) - E(Y \mid T = 0, C = 1, S = 0) \\ &= TUT + AST(p) \end{aligned}$$

With this notation in place, we can now consider the treatment effects actually recovered by three types of experimental designs. The first of these, the blocked design, features a fixed treatment saturation in each cluster and uses within-cluster controls as counterfactuals. The second, the clustered design, features treatment saturations that equal either $\bar{p}$ or zero, meaning that treatment clusters are treated with a fixed saturation and completely untreated clusters serve as the counterfactual. The third design is the randomized saturation design; here some clusters are left untreated as controls, and then within treatment clusters the saturation of treatment is randomly assigned. In the face of the SUTVA violations modeled here, the blocked design proves highly unattractive. This is because not only are the control units subject to interference from the treatment, but because the saturation is fixed in the blocked design (typically at 50%) then we have no way of estimating the extent of this interference. To see this, note that the ITT recovered by the blocked design will be:

$$\begin{aligned} ITT^B(\bar{p}) &= E(Y \mid T = 1, S = \bar{p}) - E(Y \mid T = 0, S = \bar{p}) \\ &= c(TUT + AST(\bar{p}) - ASNT(\bar{p})) \end{aligned}$$

whereas the clustered design recovers the correct ITT as long as SUTVA holds between clusters:

$$ITT^C(\bar{p}) = E(Y \mid T = 1, S = \bar{p}) - E(Y \mid T = 0, S = 0)$$

$$= c(TUT + AST(\bar{p})) + (1 - c)(ASNT(\bar{p}))$$

The bias in the blocked estimate of the ITT is given by:

$$Bias^B(\bar{p}) = E(Y \mid T = 0, S = \bar{p}) - E(Y \mid T = 0, S = 0)$$

$$= ASNT(\bar{p}).$$

This means that the blocked design is thrown off by the expected value of the spillovers in the counterfactual, a term which is in turn determined by the saturation of the blocked design $\bar{p}$. A clustered design in which $\bar{p} < 1$ is a a partial population experiment, and such designs can recover an unbiased estimate of the Average Spillover effect among the Non-Treated via

$$ASNT^C(\bar{p}) = E(Y \mid T = 0, S = \bar{p}) - E(Y \mid T = 0, S = 0)$$

or by taking the difference between the within-village controls and the pure controls. What a clustered design is unable to do to estimate how any of the treatment or spillover effects vary with saturations (because these are fixed in a typical clustered design), and consequently it also cannot differentiate $AST(p)$ from $ASNT(p)$, or $ASNT(p)$ from TUT. Further, we cannot estimate $TOT(p)$, because without maintaining SUTVA within-cluster we cannot use the clustered design to undertake the typical strategy of recovering ToT effects by instrumenting for uptake of treatment with offering of treatment. Put differently, instrumenting for uptake with offering imposes SUTVA within clusters as an exclusion restriction and so $ToT(\bar{p}) \neq \frac{ITE(\bar{p})}{c}$ unless $ASNT(\bar{p}) = 0$.

The Randomized Saturation design, by contrast, directly randomizes S_j at the cluster level and thereby provides experimental identification as to the ways in which the intensity of treatment drives outcomes. One can either draw treatment village saturations continuously or pick a discrete set of saturations, but in either case a point mass should be placed at zero to provide sufficient power to identify the pure control outcome.

With this randomized variation in saturations, we can now estimate all parameters of interest. To motivate how spillover effects differ across groups, Figure 1 provides an example of a positive spillover and illustrates how the $ITE(p)$, the $ASNT(p)$, and the $TOT(p)$ will vary across the distribution of treatment saturations. In terms of estimation, we can look across the conditional

distribution of to estimate saturation-dependent Intention to Treat Effects non-parametrically:

$$ITT^{RS}(p) = E(Y \mid T = 1, S = p > 0) - E(Y \mid p = 0),$$

and similarly for the Average Spillover on the Non-Treated:

$$ASNT^{RS}(p) = E(Y \mid T = 0, S = p > 0) - E(Y \mid p = 0).$$

We can then exploit this direct estimate of the saturation-contingent spillover effect among non-treated units to recover an estimate of the Treatment on the Treated for each value of p:

$$ToT^{RS}(p) = \frac{\hat{ITT}^{RS}(p) - (1-c)\hat{ASNT}^{RS}(p)}{c}$$

Using the entire sample, we can estimate the usual experimental intention-to-treat by defining a dummy $W_{ij} = 1(T_{ij} = 0, S_j > 0)$ for the within-cluster controls, and running the regression:

$$Y_{ij} = \eta_0 + T_{ij}\eta_1 + W_{ij}\eta_2 + X_{ijk}\gamma_k + \epsilon_{ij} \tag{1}$$

Given the variation in p this now becomes an expectation over the empirical distribution of saturations, so $E(ITT(p)) = \eta_1$ and $E(ASNT(p)) = \eta_2$. Because of the maintained assumption of homogeneous spillovers for all untreated eligibles, we can back out $E(ToT(p)) = \frac{(\eta_1 - (1-c)\eta_2)}{c}$.

A parsimonious description of the saturation effects can be obtained through:

$$Y_{ij} = \beta_0 + T_{ij}\beta_1 + W_{ij}\beta_2 + (T_{ij} * S_j)\beta_3 + (W_{ij} * S_j)\beta_4 + X_{ijk}\gamma_k + \epsilon_{ij} \tag{2}$$

In this regression β_3 gives the linearized slope of $\frac{d(ITT(p))}{dp}$ and β_4 gives $\frac{d(ASNT(p))}{dp}$, so we can estimate $\frac{d(AST(p))}{dp} = \frac{\beta_3 - (1-c)\beta_4}{c}$, and a test for $\frac{d(AST(p))}{dp} = \frac{d(ASNT(p))}{dp}$ is given by an F-test of the hypothesis that $\beta_3 = \beta_4$. Because the coefficients β_1 and β_2 become intercepts estimated where $p = 0$, $TUT = \frac{\beta_1}{c}$ and β_2 functions like a hypothesis test for the linearity of the spillover relationship, since there should be no spillover effect if the saturation of treatment is zero. We can combine estimates from (1) and (2) to get $E(AST(p)) = E(ToT(p)) - TUT = \frac{\eta_1 - (1-c)\eta_2 - \beta_1}{c}$. Each of the quantities given here that does not have a direct statistical test is a linear composite of scalars and other OLS

estimands, and so the hypothesis testing for these quantities is easily defined.

Instead of examining the slope of outcomes over saturations we may be interested in looking for threshold effects. This could mean testing for whether $AST(p)$ and $ASNT(p)$ have positive slope over some range of p and no relationship over others, or whether there is a discontinuous jump at some point in the distribution of the saturation. When, as in our case, the values of p were assigned discretely (rather than continuously), we have a simple test for threshold effects. We assigned saturations of 0% (meaning that only baseline dropouts and not baseline schoolgirls were treated), 33%, 67%, and 100% (meaning that we have no within-village controls). A more granular way of approaching this experiment, then, is to run the following regression:

$$Y_{ij} = \beta_0 + T_{ij}^{33}\delta^{33} + T_{ij}^{67}\delta^{67} + T_{ij}^{100}\delta^{100} + W_{ij}^{0}\sigma^{0} + W_{ij}^{33}\sigma^{33} + W_{ij}^{67}\sigma^{67} + X_{ijk}\gamma_k + \epsilon_{ij} \tag{3}$$

Here each level of saturation is entered as a separate dichotomous variable, with the counterfactual for all treatment and spillover effects being the pure control clusters. The coefficients δ and σ can then be used to form hypothesis tests for threshold effects. We can test for linearity in the underlying relationship, look for evidence of significant jumps between values of p, or test whether marginal effects are equal beyond some threshold value. Increasing the number of values that p can take will allow for more fine-grained analysis of threshold effects, but will decrease the power of each test by diminishing the number of observations available at each point. In our empirical analysis, we also modify both of these regression specifications by allowing Conditional and Unconditional transfers to have different treatment and spillover effects.

2.3 Power Comparisons across Research Designs

Assume we have a random effects data generating process and are interested in computing the MDE of the treatment effect for different randomization structures of experimental studies. The data-generating process has common cluster component of error, and is specified as:

$$\begin{aligned} y_{ic} &= \beta_0 + \beta_1 T_{ic} + v_c + w_{ic} \text{ for } i = 1, ..., n;\ c = 1, ..., C \\ w_{ic} &\sim (0, \sigma^2) \\ v_c &\sim (0, \tau^2) \end{aligned}$$

where there are C clusters with n observations in each cluster, for a total of $N = nJ$ observations in the whole sample. Let $u_{ic} = v_c + w_{ic}$ be the total error, which has variance $\tau^2 + \sigma^2$. Random effects assumes $Cor(v_c, T_{ic}) = 0$ which we have since T_{ic} is randomly assigned. Errors in a random effects model are homoskedastic.

Let there be a two-stage randomization procedure for the field experiment: stage 1 is a randomization at the cluster level, in which each cluster is assigned a saturation $T_c \in \Pi$ and stage 2 is a randomization at the individual level, where each individual within a cluster is assigned treatment status $T_{ic} \in \{0, 1\}$ according to the realized saturation level of the cluster in stage 1. Let $f(\pi)$ represent the stage 1 distribution over possible treatment saturations for each cluster, with mean μ and variance η^2. For example, $\Pi = \{0, .33, .66, 1.0\}$ and $f(\pi) = 1/4$. Note that the outcome of the stage 1 randomization specifies the distribution of treatment status for stage 2:

$$P(T_{ic} = 1 | T_c = \pi) = \pi$$

A two-stage randomization procedure is completely specified by the pair $\{\Pi, f(\pi)\}$. Assume all clusters have the same treatment probability (i.e. $f(\pi)$ is the same for all clusters).

2.3.1 Using within-cluster controls as counterfactuals

The treatment probability for each girl is:

$$\begin{aligned} E[T_{ic}] &= \sum_{\Pi} P(T_{ic} = 1 | T_c = \pi) P(T_c = \pi) \\ &= \sum_{\Pi} \pi f(\pi) = \mu \end{aligned}$$

and the variance in treatment status is

$$Var[T_{ic}] = \mu(1 - \mu)$$

since $E[T_{ic}^2] = E[T_{ic}]$. The correlation of treatment status between two girls in the same cluster is:

$$\rho = \frac{E[T_{ic}T_{jc}] - E[T_{ic}]E[T_{ic}]}{Var[T_{ic}]} = \frac{\eta^2}{\mu(1 - \mu)}$$

(see proof in the Appendix for the math). This is the key expectation that allows for two components to the randomization: at the cluster and individual level. This two-stage randomization introduces correlation between the unconditional treatment status of two girls in the same cluster. Conditional on the treatment status of the cluster, this correlation is once again zero. Note $E[T_{ic}T_{jd}] = E[T_{ic}]E[T_{id}]$ for different clusters, since the treatment status of individuals across clusters is independent, and thus the correlation between two girls in different clusters is zero.

The following theorem characterizes the standard error of the average treatment effect, $SE\left(\widehat{\beta}_1\right)$, for this structure randomization

Theorem 1 *Suppose a two-stage randomization procedure* $\{\Pi, f(\pi)\}$ *is used to determine treatment status and within-cluster controls are used as counterfactuals. First, each cluster is assigned a saturation* $T_c \in \Pi \subset [0,1]$ *according to the distribution* $\pi \sim \left(\mu, \eta^2\right)$*. Second, each individual in the cluster is assigned a treatment status* $T_{ic} \in \{0,1\}$ *according to the distribution* $P(T_{ic} = 1|T_c = \pi) = \pi$*. Then the standard error of the estimated treatment effect is*

$$SE\left(\widehat{\beta}_1\right) = \frac{1}{nC} * \left[\left(\frac{1+\rho(n-1)}{\mu(1-\mu)}\right)\tau^2 + \left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right]$$

where ρ *is the unconditional correlation of treatment status between two girls in the same cluster and can be defined in terms of the distribution of cluster-level treatment saturations:*

$$\rho = \frac{\eta^2}{\mu(1-\mu)}$$

Proof. See Appendix. Note this expression is taking $n \to \infty$ and $C \to \infty$. For finite n, this is not an accurate expression for the standard error of $\widehat{\beta}_1$. ■

The first term in the brackets captures the variation in $\widehat{\beta}_1$ due to cluster-level variation, and the second term captures the variation in $\widehat{\beta}_1$ due to individual variation.

This expression can be used to calculate the Minimum Detectable Effect, the smallest value of β such that it is possible to distinguish from 0 when β is the true value of the underlying parameter (i.e. reject null hypothesis $\beta = 0$) . The null hypothesis is rejected with probability k (power) when

significance level α is used to compute the test statistic.

$$MDE = [t_{1-k} + t_{\alpha}] * \sqrt{SE\left(\widehat{\beta}\right)}$$

Fixing μ, this expression is minimized at $\eta^2 = 0$. Thus, the randomization procedure that minimizes the variance of cluster level saturations leads smallest MDE, for any intended treatment saturation μ. Introducing variation in the treatment saturation of clusters results in a power loss. Additionally, for any variance, $\mu = 0.5$ minimizes $SE\left(\widehat{\beta}_1\right)$.

Corollary 2 *The optimal randomization procedure is as follows:*

- *Fix any overall sample treatment saturation, μ. The smallest MDE occurs when there is no variation in the treatment saturations across clusters. This is known as the blocked randomization design.*
- *The μ that minimizes the MDE is $\mu = 0.5$.*

Thus, the optimal randomization procedure is $\{\Pi = \{0.5\}, f(\pi) = 1_{\pi=0.5}\}$, which is the blocked randomization design with a 50% saturation.

Corollary 3 *Suppose the underlying data-generating process is iid observations so that $\tau^2 = 0$. Fix μ, the mean of the assigned saturation in the first stage of the randomization procedure. Then for any support Π and variance η^2, the standard error of the treatment effect, $SE\left(\widehat{\beta}_1\right)$, is constant. Therefore, the randomization structure is irrelevant.*

Proof. *This follows directly from the expression for $SE\left(\widehat{\beta}_1\right)$. When $\tau^2 = 0$, $SE\left(\widehat{\beta}_1\right)$ is independent of η^2.* ■

Clustered design:

Suppose each cluster is treated with probability P, and individuals in each cluster are treated with probability 1, conditional on the cluster being treated.

Plugging these in to the theorem yields:

$$SE\left(\widehat{\beta}_1\right) = \frac{1}{nC} * \left(\frac{n\tau^2 + \sigma^2}{P(1-P)}\right)$$

which is also as it should be.

Blocked design:

Suppose each cluster is treated with probability 1, and individuals in each cluster are treated with probability P, conditional on the cluster being treated.

Plugging these in to the theorem yields:

$$SE\left(\widehat{\beta}_1\right) = \frac{1}{nC} * \left(\frac{\tau^2 + \sigma^2}{P(1-P)}\right)$$

Taking the sample size $N \to \infty$ can be achieved by either taking $n \to \infty$ or $C \to \infty$. These yield different results for the distribution of $\widehat{\beta}_1$. Note that in the limit, as $n \to \infty$, the blocked design is equivalent to individual level randomization. For a blocked design, the saturation of each cluster converges to P as $n \to \infty$, and for individual level randomization, the saturation of the sample (and therefore each block) converges to P as $n \to \infty$. This holds for a fixed C. However, for finite n, this is not the case: the variance of the saturation of each cluster will be lower with the blocked design than with individual randomization. In both cases, the saturation of the overall sample converges to P as $C \to \infty$.

Randomized Saturation design:

Take a randomized saturation design in which the saturations have a discrete support and there is an equal assignment probability of each saturation.

Corollary 4 *Suppose that we compare the three designs holding $\mu = P$ constant across all three. Then, the Randomized Saturation design will have a power that is intermediate between the Clustered and Blocked designs.*

Proof. *With the population treatment saturations comparable across designs the denominator of the expression for $SE\left(\widehat{\beta}_1\right)$ is the same in all designs. The differences occur solely in the term multiplied times τ^2 in the numerator, and since $1 < [1 + \rho(n-1)] < n$, then $MDE^B < MDE^{RS} < MDE^C$.* ■

Partial-Population design

It is also interesting to use this framework to examine a partial population experiment in which a fixed fraction of the sample is treated in a subset of the clusters, and then to use this benchmark

to examine how varying the treatment saturations from this fixed value will alter power. Suppose 50% of clusters are treated at 50% and the remaining 50% of clusters are pure controls

Holding μ constant, note that the impact of randomizing saturation in treated clusters affects $SE\left(\widehat{\beta}_1\right)$ solely through the variation η^2, which is the numerator of ρ (Recall

$$\begin{aligned} SE\left(\widehat{\beta}_1\right) &= \frac{1}{nC\mu\left(1-\mu\right)} * \left[\left(1+\rho\left(n-1\right)\right)\tau^2+\sigma^2\right] \\ \rho &= \frac{\eta^2}{\mu(1-\mu)} \end{aligned}$$

And thus any move in the direction of randomizing saturations results in a power loss.

2.3.2 Using only pure controls as counterfactuals:

Varying the treatment saturation within treated clusters varies the number of treated girls per cluster. When analyzing the treatment effect by comparing treated girls to pure controls, this is equivalent to an experiment where the number of girls in each cluster varies for treatment clusters (since the within cluster controls are discarded). Let n_c be the number of girls in cluster c and $N = \sum_{c=1}^{C} n_c$ be the total number of girls.

$$SE\left(\widehat{\beta}_1\right) = \frac{1}{N} * \left[\left(\frac{\left(1+\rho\left(\phi-1\right)\right)}{\mu\left(1-\mu\right)}\right)\tau^2 + \left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right]$$

Where

$$\phi = \frac{\sum_{c=1}^{C} n_c^2}{N}$$

is the average of n_c^2. Note ϕ is minimized at $n_c = n$ for all c, and is equal to $\phi = n$. Note that it now may be optimal to have $\mu > 1/2$, as this could potentially offset some of the power loss from varying n_c among the treated clusters. (ϕ depends on μ)

A further important issue that relates to the power of the randomized design is the manner in which the saturations themselves are selected. In our study we chose them from four discrete values (0, 33, 67, and 100 percent). It would also be feasible to pick the saturations from a continuous distribution; each strategy has a specific advantage. The benefit of picking from a discrete distribution is that it maintains the power to test for impacts at each saturation relative to the pure control,

and provides a relatively powerful test for whether impacts differ across any two specific saturation levels. A continuous distribution, on the other hand, would in principle permit the detection of the exact location of a saturation threshold in the data, while the discrete version would only tell us which two points that threshold between. If the true relationship with the saturation is linear then the two designs should be equivalent.

2.4 Summary: Tradeoffs in the Use of the Randomized Saturation design

This section has formalized the benefits and disadvantages of randomizing the cluster-level saturation of treatment. Perhaps the most natural contrasting design is a partial population experiment in which half of the sample in half of the clusters are treated and the remainder of clusters serve as pure controls. The randomization saturation opens up empirical investigation of additional questions relative to the partial population design. First, we have randomized variation in intensity of treatment and therefore can look for the existence of saturation or threshold effects whereby outcomes among treated or untreated agents display interesting variation across the distribution of treatment intensity. Secondly, because we can measure spillovers among untreated agents we can impose functional form and make an assumption about the homogeneity of spillovers for the untreated and thereby back out a very rich set of parameters about how the treatment on the treated varies from the person uniquely treated in a cluster all the way across the distribution of the saturation to universal treatment.

The cost of randomizing saturations is power. As Section 2.3 demonstrates, blocked designs maximize power for any overall treatment probability but this design is biased in the face of SUTVA violations. Moving from a partial population to a randomized saturation design can be thought of as shifting from a blocked design within the subset of clusters that receive treatment towards one that varies saturation, and hence there is a power hit. The decrease in power comes from the clustered effect in the residual, which is innate, and the variation in the treatment saturation itself. Hence, the power loss is an inescapable consequence of the effort to create variation in saturations. By providing closed-form solutions for power calculations in multi-level experiments we hope to be able to assist researchers in thinking through the contexts in which they have a sufficient number of clusters to justify taking the risk on randomizing treatment intensity. We return to the issue raised at the beginning of the chapter to note that while the power loss from the randomized saturation

design will be larger if the intra-cluster correlation is higher, the presence of such a correlated effect is the entire justification for the desire to study spillovers in the first place.

3 Data and Design

3.1 Data and Sampling.

The empirical exercise on which our analysis is based was conducted in Zomba District, Malawi. We first sampled from Enumeration Areas (EAs, equivalent to a few villages) in a manner that underrepresented wealthy, urban areas (where we expected the economic impacts of the program to be muted) and distant rural areas (that were too costly to treat on a regular basis). Within sampled EAs, we then randomly sampled never-married 13-22 year old girls, oversampling older girls and girls who had already dropped out of school, and we conduct the analysis using weights to make the results representative for the average eligible girls in study villages. 88 EAs were assigned to serve as treatments, and 88 as controls.

Dropouts at baseline were always sampled and always given conditional cash transfers (CCTs) while additional experiments were conducted within the larger sample of girls who were in school at the time the program began. First, as described in Baird et al. (2011), in 46 EAs the baseline schoolgirls were given CCTs, while in 27 EAs this stratum instead received unconditional transfers (UCTs). The unique feature of the experiment described in this paper is that the EA-level saturation of treatment within the baseline schoolgirl sample was also directly randomized into four bins: in 15 EAs none of them received treatment, in 24 one third did, in 25 two-thirds did, and in 24 all did. Figure 2 presents a schematic of the randomized saturation design used here, and Figure 3 gives a map of Zomba district and the location of the CCT treatment, UCT treatment, and control EAs.

Treatment began at the end of 2007, and continued for two years. In the CCT arm, girls were given transfers of between $5 *and* 15$ per month only if they continued attending school 80% of the time or more. Those who stopped attending school for more than one consecutive month would have their payments stopped, but reinstatement of transfers was always permitted for treatments who later returned to school. The UCT arm featured the same transfer system, but in this treatment the transfers were received regardless of whether the girl attended school, a point that was reinforced at the time of monthly cash transfers.

We use data from several sources. The annual SIHR Household Survey consists of a multi-topic questionnaire administered to the households in which the selected sample respondents reside. The survey consists of two parts: one that is administered to the head of the household and another that is administered to the core respondent, i.e. the sampled girl from our target population. The former collects information on the household roster, dwelling characteristics, household assets and durables, and consumption. The core respondent survey provides information about her family background, her education and labor market participation, her health, her dating patterns, sexual behavior, marital expectations, knowledge of HIV/AIDS, her social networks, as well as her own consumption of girl-specific goods (such as soaps, mobile phone airtime, clothing, braids, sodas and alcoholic drinks, etc.). This paper utilizes data from the baseline survey (October 2007-February 2008) to perform balance checks, and data from the two follow-up waves (October 2008-February 2009 and February-June 2010) to analyze impacts.

We conducted three additional forms of data collection to provide independent outcome measures over critical variables. First, we conducted a School Survey that visited the schools of all study subjects who reported being enrolled as of the round 2 survey and collected data on enrollment and attendance directly from the school itself. Baird and Ozler (2011) demonstrate that this independently collected data is more reliable than the self-reports, and so in studying enrollment we use the independent school reports. While a school survey was conducted in round 3, this was done only for a sample of respondents, and hence in this analysis we use the round 2 data on number of terms attended, a variable which goes from a minimum of zero to a maximum of three.

To gain a rigorous measure of learning rather than simply school enrollment, we administered independent tests for English, mathematics, and cognitive skills to subjects in their homes at the time of the round 3 survey. The tests were developed by a team of experts at the Human Sciences Research Council according to the Malawian curricula for these subjects for Standards 5-8 and Forms 1-2. In addition, to measure cognitive skills, we utilized a version of Raven's Colored Progressive Matrices that was used in the Indonesia Family Life Survey (IFLS-2). These tests were administered by trained proctors at the residences of the study participants and were always administered after the household survey, but never on the same day. The order of the math and English tests were randomized at the individual level and the Raven's test was always administered last.

Finally, we conducted biomarker testing for HIV/AIDS, herpes simplex-2, and syphilis. This

testing was done only in a sample of EAs during round 2 and no within-village controls were included at that time. Round 3 biomarker testing was done for the full study sample, but was conducted six months after the program had ended and thus serves as a somewhat unattractive compromise between the concentrated treatment effect provided in Baird et al. (2011) and a long-term outcome, for which testing will commence during 2012. Nonetheless the need for a full sample requires us to use the round 3 data for this exercise.

The outcomes used in the empirical analysis, then, are the round 2 school survey data on enrollment, round 3 self-reported data on marriage pregnancy, round 3 data on the total expenditures on girl-specific consumption for the core respondent as well as the total amount spent by the core respondent on girl-specific consumption of other girls, round 3 testing data on English and cognitive test scores, and round 3 biomarker data on HIV/AIDS.

The analysis is conducted among the 2,907 sampled baseline schoolgirls, pooling together the treatment, spillover, and pure control samples. Table 1 provides summary statistics for two sets of outcomes; in Panel A we see baseline covariates, and in Panel B we present statistics on the endline (round 3) dependent variables used in the analysis.

3.2 Research Design

We conducted a randomized saturation within the sample of baseline schoolgirls. The fact that the saturations were done within the sample causes the actual saturation of treatment to be lower than intended by the inverse of the sampling ratio in the group. Figure 4 shows the distribution of the observed saturations by those randomly assigned, first showing the fraction within the sample and then the fraction within the population. Overall, the saturations are similar; the average weight in the overall schoolgirl sample is only 1.4 and the maximum weight is 7.3, indicating that in reality the observed true EA-level saturations are close to those assigned. The correlation between the assigned and actual saturations at the EA level is .86, while the assigned saturations are completely orthogonal to the sampling weights in the village, with a correlation of .03. We can instrument for actual saturations with intended, or simply use the intended saturations directly; the answers are similar subject to a scaling factor whereby the instrumented impacts are increased by a factor equal to the average sampling weight.

Table 2 shows balance tests, focusing both on the balance of the treatment/within-village control

split as well as the balance of the randomized saturations across EAs. We use a structure that mimics the regressions to be used in the next section, conducting cross-sectional comparisons at the individual level while clustering standard errors at the EA level to account for the design effect. The first two rows of results show the balance of the treatment as compared to the pure controls, and replicate the balance statistics shown in Baird et al. (2011a): there is some imbalance on baseline age (CCT group) and asset index (UCT group), but otherwise the experiment looks well-balanced.

In the next two rows we show the balance of the within-village controls as compared to the pure controls (with the point estimates in these rows pertaining to the 0% saturations because of the inclusion of saturations as well); here we see absolutely no evidence of imbalance. We then further examine the balance of the saturations themselves, looking at whether there is a slope to the outcomes in the spillover sample relative to the control as the EA-level treatment saturations increase. We see one significant imbalance out of 20 comparisons, exactly what we would expect from random chance, and hence the balance of the saturations also appears clean.

One natural question that arises in the analysis of such a multi-dimensional experiment is whether we are sufficiently powered to find meaningful spillover effects at all. The sample sizes for many of these comparisons is relatively small, as can be seen from Figure 2. One simple way of addressing this question is to exploit the standard errors of the point estimates in our balance tests. Of course, multiplying each of these SEs by 1.96 gives us the minimum detectable effect given the variance structure observed at baseline, and so the ratios of the SEs in the spillover groups relative to the treatment groups gives us a simple way to infer the relative power of the comparisons. If we run balance tests using only the CCT/UCT dummies and the within-village CCT/UCT dummies and then take the average ratio of the SEs across all outcomes for the treatment versus within-village controls, we find that the CCT spillover sample actually has higher power than the treatment sample (ratio = 1.13) while the power of the UCT spillover sample is only marginally lower than the treatment group (ratio = 0.82). Hence while lack of power may be something of a concern in the UCT spillover group, overall the study is just as adequately powered to detect spillover effects as it is to detect direct treatment effects.

4 Analysis at the Cluster Level

4.1 Linear Analysis of Saturation Effects

In this section, we present results on the spillover and threshold results that flow directly from the research design. We begin with the simplest estimation Equation (1), examining all baseline schoolgirls in the last survey wave and including dummy variables for being in the treatment sample, the spillover sample, and a set of covariates, and standard errors clustered at the EA level to reflect the design effect in the experiment. Table 3 presents two sets of results for each outcome, first showing simple ITT and ASNT effects by estimating Equation (1) in the odd-numbered columns, and then proceeding to test for the presence of saturation effects using Equation (2) in the even-numbered columns. The treatment effects of the program have been document elsewhere and so we focus here on the investigation of the spillovers. Looking first at the average spillover effect on within-village controls, we see no significant ASNT effect on any outcome when we examine this group as a whole. This is the first evidence that the strong spillover effects detected in many other analyses of CCTs are not present here, in a study that was specifically geared to look for them.

In the even-numbered columns we examine the slope of outcomes over the randomized distribution of the treatment saturation. It is possible that there would be strong saturation and threshold effects even for an outcome where spillovers do not exist on average for the within-village controls. Here there are results that are significant at the 10% level on enrollment and on expenditures. The improvements in enrollment seen as a result of treatment are lower for EAs that see higher treatment saturations, suggesting that the improvements engendered by cash transfer programs are subject to mild congestion effects as we move towards having 100% of girls treated.[3] Strong evidence of monetary spillovers is provided in columns between eligible girls is provided in column 16, which shows that when the saturation of treatment in the village rises, within-village controls see greater expenditures on their personal consumption by others.[4]

The bottom five rows of Table 3 explicitly calculate each of the relevant treatment effects. First note that the regression coefficients on the treatment saturations give the linearized slopes of the outcomes with respect to treatment saturations. The $E(ITT)$ is the average intention to treat

[3]Recall that because the average sampling weight among baseline schoolgirls is 1.4, that 100% saturation of the study sample equates to roughly 71% effective treatment saturation in the entire population of eligibles.

[4]These items are soap, mobile phones and airtime cards, girls shoes and clothing, makeup and hairdressing, snacks, drinks and food consumed outside the home, birth control, handbags, and transport costs.

across the empirical distribution of observed saturations, given by the first row in the odd-numbered columns. The $E(ASNT)$ is the comparable quantity for the within-village controls. We can invoke the assumptions laid out in Section 2 to interpret the first row in the odd-numbered columns as the intercept term across the saturation distribution, and so it gives the TUT. We can then use the average R3 compliance rate across the CCT and UCT of .75 to calculate $E(ToT)$, and given these last two quantities we then have $E(AST) = E(ToT) - TUT$.

We can frame this lack of spillovers in a positive context by pointing out that there are at least two negative spillovers that the literature might lead us to expect that we do not find here. The first of these is a congestion effect in schooling that could arise as the program increases enrollment and hence class size. The fact that the enrollment rates and learning outcomes of the within-village controls do not fall, and cognitive functioning may even improve, is encouraging in this context. Then, on the marriage and pregnancy results; the significant findings on HIV and HSV-2 that were reported in Baird et al. (2011b) could be entirely illusory if it were the case that the program was merely diverting a fixed amount of sexual activity from the girls now receiving transfers to their contemporaries who were not. In this case, we could find a significant decrease in biomarker prevalence comparing the treated to the pure controls, but a counterbalancing increase among the within-village controls would imply that there is in fact no real impact on treatment communities. For each of marriage, pregnancy, and HIV status, the coefficient on treatment saturations for the within-village controls is actually negative. The absence of any evidence of deleterious spillover effects on these outcomes indicate that this improvements in marriage, pregnancy, and infection rates induced by the treatment are real and are not simply displacing sexual activity.

Baird et al. 2011 show that the treatment effects of the program were divergent across the conditional and unconditional arms, with the CCT tending to improve schooling and human capital acquisition and the UCT proving superior in terms of decreasing marriage and pregnancy. In Table 4, then, we split the two treatment arms, and because these quantities were randomized at the EA level we can also clearly conceptually split the spillover effects between those observed in EAs where the treatment is conditional versus those where it is unconditional. For EAs offered a CCT define $T_{ij}^C = 1$ and $W_{ij}^C = 1$ and for UCTs $T_{ij}^U = 1$ and $W_{ij}^C = 1$, with $T_{ij} = T_{ij}^C + T_{ij}^U$ and $W_{ij} = W_{ij}^C + W_{ij}^U$. We can then estimate the 'split' counterparts to (1) and (2), namely

$$Y_{ij} = \eta_0 + T_{ij}^C \eta_1^c + T_{ij}^U \eta_1^U + W_{ij}^C \eta_2^c + W_{ij}^U \eta_2^U + X_{ijk}\gamma_k + \epsilon_{ij} \quad (4)$$

and

$$Y_{ij} = \beta_0 + T_{ij}^C \beta_1^C + T_{ij}^U \beta_1^U + W_{ij}^C \beta_2^C + W_{ij}^U \beta_2^U + (T_{ij}^C * S_j)\beta_3^C + (T_{ij}^U * S_j)\beta_3^U + (W_{ij}^C * S_j)\beta_4^C + (W_{ij}^U * S_j)\beta_4^U + X_{ijk}\gamma_k + \epsilon_{ij} \quad (5)$$

Again, the odd-numbered columns of Table 4 give estimates of simple intention-to-treat effects using (4) and the even-numbered columns provide estimates of saturation effects using (5). Splitting apart the two treatments we now see the treatment effects much more distinctly, but again the overall spillover effect are quite muted even when examined in a manner that shows much starker treatment effects. The sole $E(ASNT)$ that proves to be significant using (1') is the cognitive functioning of the CCT within-village controls; here we see very stark improvements in treatment girls but improvements among untreated girls in the same villages that are if anything even larger than the raw treatment effects. This provides some tantalizing evidence that meaningful improvements in learning can be achieved as an environmental effect of the emphasis on schooling that is brought through conditional cash transfers.

When we move to examining saturation effects through (5) in the even-numbered columns, the results are consistent with the effects found to be significant so far. Despite the strong increases in enrollment and test scores seen as a result of the CCT treatment, all three outcomes have negative slopes on saturations with the slope significant at the 10% level for English achievement. This again suggests that there is some form of congestion effect that is causing the educational improvements of the CCT treatment to decline as the fraction of the eligible sample treated increases. Conversely, we again see evidence that educational outcomes for the within-village controls, particularly cognitive test scores, improve with the saturation of treatment. Intriguingly, this indicates that $\frac{d(ASNT(p))}{dp} > 0$ while $\frac{d(AST(p))}{dp} < 0$, suggesting a welfare tradeoff between treated and untreated units as the saturation increases. Despite the strong effects of the UCT on marriage, pregnancy, and HIV, no significant spillovers of either treatment are found on these outcomes. Expenditures on girl-specific items display upward-sloping saturation effects under both treatments, with CCT saturations increasing the amount girls spend on themselves and UCT saturation increasing the amount that

others spend on them.

4.2 Granular Analysis of Saturation Effects

We can relax the assumption of linearity in the treatment saturations and push the data to the maximum granularity permitted by our research design by using dichotomous variables for each of our treatment saturations. This is the specification that permits us to examine threshold effects directly, and by dichotomizing saturations for treated and within-village controls separately, we can test for distinct threshold effects within each of these groups. Table 5 estimates Equation (3), and these cell-specific treatment effects permit us to observe directly the variation across saturation levels that identified the linearizations estimated in the even-numbered columns of Table 3.

The data can be pushed to one further level of granularity by estimating a variant of Equation (3) that splits CCT and UCT EAs:

$$\begin{aligned} \mathrm{Y}_{ij} = \beta_0 &+ T_{ij}^{C33}\delta^{C33} + T_{ij}^{C67}\delta^{C67} + T_{ij}^{C100}\delta^{C100} + T_{ij}^{U33}\delta^{U33} + T_{ij}^{U67}\delta^{U67} + T_{ij}^{U100}\delta^{U100} \\ &+ W_{ij}^{C0}\sigma^{C0} + W_{ij}^{C33}\sigma^{C33} + W_{ij}^{C67}\sigma^{C67} + W_{ij}^{U0}\sigma^{U0} + W_{ij}^{U33}\sigma^{U33} + W_{ij}^{U67}\sigma^{U67} + X_{ijk}\gamma_k + \epsilon_{ij} \quad (6) \end{aligned}$$

In Table 6 each level of saturation and each treatment is entered as a separate dichotomous variable, with the counterfactual for all treatment and spillover effects being the pure control clusters. We present pictures in Figures 5-8 that provide the visual analog to these regressions, with the only difference being that Table 6 includes covariates as controls while the figures present the raw averages of outcomes within each treatment cell. Again, this table and the four pictures allow for a straightforward visualization of the data points underlying the linearized slopes estimated in Table 4. The threshold at which the enrollment impacts fall off for the treatment appear to be between 67% and 100% for the CCT, while it lies between 33% and 67% for the UCT. Cognitive test score are monotonically decreasing for the CCT treatment, and the apparent upward slope for the CCT spillover sample comes solely from a large treatment effect in the 67% saturation EAs. For marriage, pregnancy, and HIV the only significant cell-level spillover effects are negative, again confirming the absence of deleterious impacts of the program on the within-village controls. The CCT leads to large

increases in beneficiaries' expenditures on their own consumption, while others improve spending on CCT beneficiaries' consumption only when treatment saturations are high.

5 Examining Network Spillover Effects

In the previous section, we examined spillover effects as directly motivated by the cluster-level research design. We now exploit random variation at the individual level to study spillovers in two much more concentrated channels than the cluster: households and peer-group social networks. The nature of friendship networks is then examined through time (treatment effects on changes in network composition) as well as through space. We show that even if we have chosen the wrong network for our initial randomization, the intentional variation of saturation will aid in the identification of spillovers within other networks, as long as any correlation exists between the memberships of the originally selected and the actual salient network. We conclude with a direct analysis of spatial spillovers as in Miguel and Kremer 2004, and Bobba and Gignoux 2010, showing that there is no obvious contamination of our pure control.

5.1 Analysis of Spillovers within Households

We can exploit the fact that the sample contains 517 individuals that come from households with more than one girl in the study to examine spillover effects within households. To ensure that we have conditional randomness, we include a carefully constructed set of covariates. First, we control for whether the girl herself is treated or is a within-village control, to remove the direct treatment or spillover effects that were estimated in the previous section relative to the pure control. We control for the number of siblings who were baseline dropouts, because such girls were treated with probability 1 in treatment clusters. Controls are also included for the number of siblings in the sampling frame and the average sampling weight within the household, to remove endogenous variation in the number of siblings treated. Conditional on these covariates, the remaining variation in the saturation of treatment is random.

Tables 7 and 8 conduct the household-level spillover analysis, with controls included to remove endogeneity arising from the sampling probabilities of household members. When we look in this concentrated network where motivational effects are likely to dominate, treatment exerts a positive

spillover effect on enrollment. The sisters of study girls are more likely to attend school if their sisters are treated, an effect whose magnitude is similar for girls who are themselves treated and untreated, although the effect is only significant for the treated. While having one girl attend school may ease the way for others in respects such as transportation to school, it is equally plausible that the need for adolescent labor in the household would have resulted in a particularly strong negative spillover on female siblings, and so these positive coefficients are encouraging. Table 7 shows the untreated sisters of UCT-treated girls performing worse in cognitive tests, however.

Households appear to generate even stronger beneficial spillovers of the program on marriage, pregnancy, and HIV. Table 8 shows that when two sisters receive CCT treatment together, marriage decreases by 6% and pregnancy by 5%. Untreated girls see their HIV rates fall by 4% when their sisters are treated with either a conditional or unconditional transfers, and all the coefficients for marriage and pregnancy within these cells are negative as well. Surprisingly, the within-household spillovers on expenditures are very limited. While the signs in Table 8 are all as we might expect (treatment increases the amount spent on consumption of treated and untreated girls, increases the amount that treated girls spend on others, and decreases the amount that untreated girls spend on others), they are all insignificant. In the split analysis of Table 9 the only significant effect is that untreated sisters of CCT girls spend less on the consumption of other girls. Hence, while we might expect the expenditure spillover effects to be concentrated between sisters within households, this appears not to be the case.

5.2 Analysis of Spillovers within Social Networks

As a part of the baseline and Round 2 surveys, we asked girls to list their five closest friends, as well as providing basic information on these friends such as their names, village of residence, schooling status, age, and religion. Using these covariates we try to find these friends in our own sample, a linking that generates a rich set of covariates on the attributes of the friends, including their treatment status. On average girls list 4.35 friends, of whom 3.12 would be in the sampling frame for the study. Given an average sampling rate of 72.5%, if we correctly found the friend in our sample in every case we should observe an average of 2.26 friends. in the study of Just under 20% of the friends listed could be found in the sample overall, and another 23% linked to girls in the study (i.e. the friend herself wasn't found, but at least two girls in the study listed her). For 48% of girls

none of their friends were found the survey, for 30% 1 friend is found, for 14% we found 2 friends, for 6% we found 3, and for 1.5% we found four or more of their friends in our sample.

The covariates included in the analysis of social networks must reflect the fact that whereas we always knew if two study girls were siblings, in many cases we failed to link two girls who were in fact friends. Because we only observe treatment status for the girls we can find, the endogenous variation in our ability to find friends in the sample enters into our estimate of the number of friends treated. To account for this, we include the same controls as in the household analysis, and then dummy out the distribution of the number of friends found in the survey. This analysis proves to be quite robust to the exact functional form of the control variables.

Table 9 estimates spillover effects using variation in treatment status of the friends. The results of this analysis are broadly consonant with the EA-level regressions. The most important difference from the EA-level results is that within social networks we detect a positive spillover effect of treatment on enrollment. Each treated friend increases the number of terms attended in the first year for untreated girls by .11 out of a potential total of 3. This improvement is not, however, seen for treated girls. When the treatments are pooled together, there are no significant spillover effects of the treatment on achievement scores, marriage, or pregnancy. Interestingly, even in this specification we see a positive effect of the treatment on cognitive scores for within-village controls overall, indicating that this spillover is environmental and is not being transmitted through social networks.

Very importantly, when we turn to HIV we see that not only does higher saturation of treatment in the social network not expose girls to greater HIV risk, but in fact there is a significant *decrease* in prevalence as the treatment saturation increases for untreated girls. Given an overall round 3 prevalence of 6%, the marginal effect indicates that having three treated friends would completely eliminate HIV risk for an untreated girl. We consider this result to be of paramount importance, indicating that the beneficial effect of this intervention on HIV is not coming from diversion of sexual activity across subjects, and indeed seems to be casting a kind of protective halo even over girls who were not a part of the treatment. The spillover effects on expenditures are very strong, indicating that the treatment concentrates money on treated girls and saps spending that would otherwise have been directed towards the untreated.

In Table 10, following the methodology of the previous section, we now split apart the conditional

and unconditional treatment so as to be able to examine the spillover effects of the two interventions separately. Splitting the two treatments apart makes it clear that the beneficial network spillovers of the treatment on enrollment arise solely through contact with conditional girls; the coefficient for the number of unconditional friends while positive is only half as large and insignificant. The impact of the two treatments on HIV is the same in magnitude, but possibly due to the higher power of tests on the CCT program, this spillover is significant only for the conditional treatment. Expenditure spillovers again display signs that are similar across treatment arms, but somewhat surprisingly it is the unconditional treatment that seems to lead to a large contraction in girl-level consumption. This is not what we would have expected because to the extent that the treatment causes resources to be concentrated on treatment girls, this logic should only hold for the conditional treatment. At any rate there is certainly no evidence here that either treatment is improving expenditures on items specifically consumed by girls themselves.

To summarize, while we might expect social networks to display spillover effects that are both much stronger and potentially quite divergent from the spillovers observed within the diffuse network of a village, the only case in which we see distinct effects across the two analyses is for enrollment. Here, the most plausible candidate for a positive spillover effect appears to be motivational (Akerlof and Kranton 2002) while the negative effects are most likely due to congestion (Behrman, Sengupta, and Todd 2005). It is therefore quite revealing that the positive spillovers of a cash transfer program are more pronounced within social networks where the motivational effects are likely to be at a maximum, while congestion problems are spread over all those who attend the school. Similarly the strong effects of treatment within social networks on protecting girls from HIV risk are consistent with a social norm altering the behavior of those not directly treated.

5.3 Analysis of Treatment Effects on Network Composition

We now turn to an analysis of treatment effects on the dynamics of social networks. The questions about 'five closest friends' were asked again in Round 2, and this permits an analysis of the shifts in network composition that arise because of treatment. A very new empirical literature has begun to document the fact that in addition to treatment effects passing through social networks, substantial treatments may indeed alter the composition of the social networks themselves (see, for example, Bandeira et al. 2009, who find that a poverty reduction program in Bangadesh causes previously

poor recipients to form social ties to wealthier residents). We examine the change in social networks that occurred across the two rounds and test for whether these changes differ by treatment status. Table 11 reports the results of this exercise, and demonstrates clear evidence that the social networks have themselves been strongly driven by treatment.

The two core results in Table 11 are that the treatment increases the 'churn' in social networks, and that it increases the likelihood that treated girls are friends with other treated girls. Interestingly, we see that treated girls add .69 more treated friends than untreated girls do, but also lose .58 more treated friends between the two rounds than untreated girls. Thus treatment causes treated girls to both befriend and defriend other treated girls at a higher rate, but because the former effect is larger than the latter, the overall association of treated with treated increases. These results are both interesting in their own right and also make an important point for the applied conduct of network analysis in trials of consequential treatments: because social networks will realign in endogenous ways, it is critical to collect social network data at baseline in order to have clean identification of network impacts.

5.4 Robustness: Spatial Spillovers

Having examined the temporal nature of social networks, we now turn to the spatial dimension of spillover effects. Social networks, which are themselves highly spatial, appear to be important conduits for spillover effects. This raises the possibility that spillover effects might be occurring cross-cluster, violating our SUTVA assumption and causing our pure control to become contaminated. To examine this possibility, we take a very different cut on our data by examining cross-cluster spillovers using techniques based on Miguel and Kremer 2004, and Bobba and Gignoux 2010. We take GIS data on the locations of the EA centroids, and then construct cluster-level variables giving the number of EAs within 3 and 3-6 kilometers. By then separately including the number of treated clusters within these distance bands, we generate conditional randomization and can examine spillovers induced by the saturation of treatment in surrounding EAs. Any strong effects here indicate that even our pure control outcome is contaminated, and hence both our treatment and spillover estimates will be biased.

Table 12 demonstrates that this cash transfer experiment did not generate strong cross-cluster effects. There are no consistent spillovers on enrollment, sexual behavior, or HIV. We see some

borderline evidence of negative spillover effects on test scores and girl-specific consumption. If we consider treatment in adjoining EAs to induce congestion effects in local schools without creating positive motivational effects, this weak negative effects seems reasonable. To the extent that treatment induces a negative effect on test scores even in pure control villages, this indicates that the overall study may slightly overestimate the actual effect of the program on testing. There is some evidence of negative cross-cluster spillovers in consumption for girls. In contrast to Bobba and Gignoux who find large spillover effects of *Progresa* in Mexico but only on treated individuals, we find virtually no evidence that program beneficiaries experience spillovers from adjacent clusters that are any different from untreated individuals.

5.5 The Benefits of Randomized Saturations when we use Different Networks

We conclude by demonstrating a side benefit of the randomized saturation design: it increases the saturation variation present in alternate definitions of network, so long as other networks display any spatial correlation. The more closely any alternative network definition maps on to space, the more closely will the variation in saturations resemble those assigned in the experiment.

To motivate this idea with our data, we combine our actual data on social networks with a Monte Carlo exercise in which we simulate the distributions of network-level treatment saturations across different research designs. We begin from the universe of girls who (a) live in study villages and (b) were eligible for the treatment (meaning never-married 13-22-year olds) and (c) are either themselves in the study sample or are listed at baseline among the five closest friends of girls in the study sample. This provides us with a sample of 8,981 individuals in 175 EAs. We then assign placebo treatments: the Blocked treatment is assigned so that 50% of the units in each cluster are treated, the Clustered treatment is given to all units in 50% of the clusters, and the Randomized Saturation treatment is assigned at saturations of 0%, 33%, 66%, and 100% per cluster with a quarter of clusters at each saturation, so that each assignment rule results in the same overall fraction of the sample being treated (one half). Each randomization is conducted 100 times. We then map each of these placebo randomized designs back to the actual social networks to which the study subjects belong and examine the empirical distribution of treatment saturations within-network.

This combination of Monte Carlo randomization with observed social networks allows us to remove the small-sample variation that would naturally occur in any specific randomized draw

and thereby to study the limiting distributions of treatment saturations under different designs. Figure 9 provides a graphical representation of the resulting distributions, plotting the densities of treatment saturations across the three designs. First, consider the Blocked design. Not surprisingly, the treatment saturations in this design are strongly centered around 50%. This is problematic not only because there is little overall variation in the saturations (the standard deviation of the treatment saturations in the Blocked, Clustered, and Randomized Saturation designs are .287, .417, and .362, respectively) but because there are very few networks that feature no treatment to be used as pure controls. The fraction of networks with zero treatment saturation across the three designs are 9.4%, 27%, and 20%, respectively. Hence the blocked design features no networks in clusters with no treatment and fewer than 1 in 10 networks with no treatment, and thus provides a weak counterfactual for a pure control that is demonstrably free of spillovers.

The Clustered design suffers from the opposite problem; because treatment has taken place at the village level it is dominated by networks that are either entirely treated or entirely untreated. Recall than in estimating network treatment effects it is necessary to control for the treatment status of the core respondent, and so we only have statistical identification of the saturation effects when the treatment saturation of the network differs from that of the respondent. This occurs in only 41% of networks under the Clustered design but in almost 60% of networks under the Randomized Saturation design. Finally, visual inspection of Figure 9 makes it clear that the Randomized Saturation design produces an almost continuous distribution of network-level saturations, while retaining a point mass at 0%, while the Blocked and Clustered design return saturation distributions more inclined to point masses at 50%, and 0 and 100%, respectively.

The underlying phenomenon delivering this improvement in the property of the saturations in networks other than the one originally randomized on is the fact that a correlation exists between one definition of networks (social networks) and the other (location). In the limit, as these two move to being completely orthogonal then the Randomized Saturation design will do no worse and no better than the others, and as they become perfectly correlated then the second network definition will deliver saturations that look exactly like those that were randomly assigned. In this sense, randomizing saturations over spatial clusters as was done in this study is attractive for a variety of reasons. First, it is logistically relatively simple, and network membership can be easily identified ex ante without detailed fieldwork. Additionally, any straightforward implementation of

the Randomized Saturation design would require that the networks used be non-overlapping. In this exploitation of alternate candidates such as social networks, households, financial networks, or group membership is likely to present myriad problems in implementation, but because many other types of network are likely to have some spatial dimension then it is a good way of extending the power of the Randomized Saturation design into a large number of network dimensions.

6 Conclusion and Discussion

Empirical researchers have become increasingly concerned in recent years with the problem of interference between research subjects. Rigorous estimation of spillovers using experiments designed specifically for this purpose not only opens up a fascinating set of research questions free of the the reflection problem, but provides critical information for policymakers as well. Research designs that fail to anticipate spillovers correctly can be biased even with the use of a clean Randomized Controlled Trial, and results from a field study that show meaningful treatment effects but fail to observe deleterious spillovers can arrive at policy conclusions that are essentially meaningless. This paper attempts to push the frontier of field trials by discussing how novel research design can avoid pitfalls that come from violations of SUTVA and shed light on a set of critical policy parameters.

Our approach permits us to differentiate the spillovers that occur on untreated units from those that occur on treated units. In studies with non-compliance, this distinction becomes difficult to estimate because we can experimentally compare spillovers on those offered and not offered the treatment but not directly spillovers on those treated versus not treated. We present a framework suggesting that because the Randomized Saturation design generates experimental spillovers on eligible untreated units (as opposed to most extant partial population experiments) and also features variation across the full distribution of treatment saturations, it provides an attractive environment in which to back out a richer set of parameters. We invoke two critical assumptions: (1) eligible non-compliers demonstrate the same spillover effects as untreated eligibles, and (2) a functional form assumption on the nature of saturation effects (most simply, linearity). Using these two assumptions we can then nail down the Treatment Effect on the Treated as a function of the saturation intensity, and the intercept on this term at zero saturation provides the 'Treatment on the Uniquely Treated', beyond which variations in the ToT can be interpreted as spillover effects on those actually treated.

The intentional creation of variation in the intensity of treatment also substantially improves the external validity of randomized trials. Any treatment and spillover effects estimated from a partial population experiment with fixed treatment intensity cannot address how impacts would have changed had this intensity been varied. The purpose of the Randomized Saturation design presented here is directly to inject randomized variation into this intensity. While a large theoretical literature discusses the nature of potential threshold effects in network theory, public health, and education, less work has gone into thinking through how experimental design can help to shed light on these quantities. Again, clean estimation of these threshold effects is not only of deep theoretical interest, but provides critical answers to policy questions as well. For example, if we are implementing a program with fixed resources and can either treat 100% of five villages or 50% of ten villages, which will prove most cost effective? Standard research design provide little or no variation that is useful in answering this question. Further, one of the typical objections of randomized experiments is their questionable external validity with respect to a full-scale, non-experimental implementation of the program. To the extent that pushing the cluster-level saturation towards 100% is moving us from a partial equilibrium to a general equilibrium set of outcomes (because prices, norms, and congestion effects adjust more fully as saturation rises) then the Randomized Saturation design proves a richer experimental framework.

Our results help to inform the efficient design of cash transfer programs. Few of the outcomes studied here display strong responses to saturation of treatment at the village level, indicating that an additional individual treated in a new village will have roughly similar overall treatment effects to an additional individual in a currently treated village. The exception to this is expenditure on girl-related consumption, where we find a strong pattern whereby increasing the saturation of treatment redirects expenditures from treated to untreated girls, making the targeting of the program less consequential for the final pattern of expenditures. Possibly the most important result here from a policy perspective is that the beneficial impacts of our treatment on HIV reported elsewhere are not being undermined by compensatory spillovers. If anything, we find that both friends and sisters of treated girls see their HIV risks drop substantially as an indirect result of the program. This is encouraging evidence that cash transfer programs could form a meaningful part of an anti-HIV campaign in Sub-Saharan Africa, and shows how careful research design can assist policymakers in design choices.

7 References

Alix-Garcia, Jennifer, Craig McIntosh, Katherine Sims, and Jarrod Welch, "The Ecological Footprint of Poverty Alleviation: Evidence from Mexico's Oportunidades Program," Working Paper, (2011).

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

Angelucci, Manuela, Giacomo De Giorgi, Marcos Rangel, and Imran Rasul, "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment." Journal of Public Economics, 94 (2010).

Babcock, Phillip, and John Hartman, "Networks and Workouts: Treatment Size and Status-Specific Peer Effects in a Randomized Experiment," NBER Working Paper 16581.

Baird, Sarah, and Berk Özler, "Examining the Reliability of Self-Reported School Enrollment Data," Forthcoming, Journal of Development Economics, (2010).

Baird, Sarah, Craig McIntosh, and Berk Özler, "Cash or Condition: Evidence from a Cash Transfer Experiment," Forthcoming, Quarterly Journal of Economics, (2011a).

Baird, Sarah, Richard Garfein, Craig McIntosh, and Berk Özler, "Impact of Cash Transfer Program for Schooling on Prevalence of HIV and HSV-2 in Malawi: A Cluster Randomized Trial," Working Paper, (2011b).

Bandiera, Oriana, Robin Burgess, Selim Gulesci, and Imran Rasul, "Community Networks and Poverty Reduction Programmes: Evidence from Bangaldesh." LSE STICERD Research Paper No. EOPP 015 (2009).

Barham, Tania, and John Maluccio, "Eradicating Diseases: The Effect of Conditional Cash Transfers on Vaccination Coverage in Rural Nicaragua." Journal of Health Economics, 28 (2009). pp. 611-621.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh Linden, and Francisco Perez-Calle, "Conditional Cash Trasnfers in Education: Deisgn Features, Peer and Sibling Effects: Evidence from Randomized Experiment in Colombia." NBER Working Paper 13890.

Beaman, Lori, "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S.," Working paper (2010).

Behrman, Jere, Hans-Peter Kohler, and Susan C. Watkins, "Social Networks, HIV/AIDS and Risk Perceptions," PIER Working Paper 03-07, (2003).

Bobba, Matteo, and Jeremie Gignoux, "Spatial Externalities and Social Multipliers of Schooling Interventions." Working Paper.

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

Conley, Timothy and Christopher Udry, "Learning About a New Technology: Pineapple in Ghana." American Economic Review 100:1 (2010), pp. 35-69.

Drago, Francesco, and Roberto Galbiati, "Indirect Effects of a Crime Control Policy: Evidence from the Italian Prison Experiment." IZA Working Paper 5414, (2010).

Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Using Randomization in Development Economics Research: A Toolikit," CEPR Working Paper 6059. (2007).

Duflo, Esther and Emmanuel Saez, "Participation and Investment Decisions in a Retirement Plan: The Influence of Colleague's Choices." Journal of Public Economics 85 (2002), pp. 121-148.

Duflo, Esther and Emmanuel Saez, "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," Quarterly Journal of Economics, August (2003), pp. 815-842.

Gine, Xavier, and Ghazala Mansuri, "Together We Will: Evidence from a Field Experiment on Female Voter Turnout in Pakistan." Working paper, (2011).

Kuhn, Peter, Peter Kooreman, Adriaan R. Soetevent, and Arie Kapteyn, "The Effects of Lottery Prizes on Winners and their Neighbors: Evidence from the Dutch Postcode Lottery," IZA Discussion Paper 4950.

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

Macours, Karen, and Renos Vakis, "Changing Households' Investments and Aspirations through Social Interactions: Evidence froma Randomized Transfer Program in a Low-Income Country." Working Paper, (2008).

Macours, Karen, Norbert Schady, and Renos Vakis, "Can Conditional Cash Transfer Programs Compensate for Delays in Early Childhood Development?" Working paper, (2008).

Manski, Charles, "Identification of Endogenous Social Effects: The Reflection Problem," The

Review of Economic Studies, 60(3), (1993), pp. 531-542.

Manski, Charles, "Identification of Treatment Response with Social Interactions," CEMMAP Working Paper CWP01/10.

Miguel, Edward, and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," Econometrica, Vol. 72 No. 1, January (2004), pp. 159-217.

Moffit, Robert, "Policy Interventions, Low-Level Equilibria and Social Interactions," in Social Dynamics, eds. Steven Durlauf and Peyton Young, MIT Press.

Munshi, Kaivan, "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," The Quarterly Journal of Economics, May (2003), pp. 549-599.

Oster, E., and R. Thornton, "Determinants of Technology Adoption: Private Value and Peer Effects in Menstrual Cup Take-Up", NBER Working Paper No. 14828.

Rosenbaum, P., "Interference Between Units in Randomized Experiments," Journal of the American Statistical Association, 103(477), March (2007).

Rubin, D.B., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," Journal of Educational Psychology, Vol. 66 (1974), pp. 688-701.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green, "Detecting Spillover Effects: Design and Analysis of Multi-Level Experiments," Working Paper, (2011).

Tarozzi, Alessandro, Aprajit Mahajan, Brian Blackburn, Dan Kopf, Lakshmi Krishnan, and Joanne Yoong, "Micro-Loans, Insecticide-Treated Bednets and Malaria: Evidece from a Randomized Controlled Trial in Orissa," Working Paper (2011).

8 Appendix

8.1 Proof: computing a closed form expression for $SE\left(\widehat{\beta}_{OLS}\right)$ when using within-village controls

Using the underlying matrices of the model (specified in following section), we can write

$$\begin{aligned} X_c'X_c &= \begin{bmatrix} n & \sum_{i=1}^n T_{ic} \\ \sum_{i=1}^n T_{ic} & \sum_{i=1}^n T_{ic}^2 \end{bmatrix} \\ X_c'u_cu_c'X_c &= \begin{bmatrix} \left(\sum_{i=1}^n u_{ic}\right)^2 & \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n T_{ic}u_{ic}\right) \\ \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n T_{ic}u_{ic}\right) & \left(\sum_{i=1}^n T_{ic}u_{ic}\right)^2 \end{bmatrix} \end{aligned}$$

Note the following:

- $T_{ic}^2 = T_{ic}$
- T_{ic} is independent of u_{ic} so $E\left[f(u_{ic})g(T_{ic})\right] = E\left[f(u_{ic})\right] * E\left[g(T_{ic})\right]$
- $E\left[u_{ic}^2\right] = \tau^2 + \sigma^2$
- $E\left[u_{ic}u_{jc}\right] = \tau^2$ if $i \neq j$ which is $Cov(u_{ic}u_{jc})$
- $E\left[u_{ic}u_{jd}\right] = 0$ if $c \neq d$
- $E[T_{ic}] = \sum_{\Pi} P(T_{ic} = 1|T_c = \pi)P(T_c = \pi) = \sum_{\Pi} \pi f\left(\pi\right) = E[\pi] = \mu$
- $E\left[T_{ic}T_{jc}\right] = \lambda$

$$\begin{aligned} E\left[T_{ic}T_{jc}\right] &= \sum_{\Pi} P(T_{ic} = 1, T_{jc} = 1, T_c = \pi) \\ &= \sum_{\Pi} P(T_{ic} = 1|T_{jc} = 1, T_c = \pi)P(T_{jc} = 1, T_c = \pi) \\ &= \sum_{\Pi} P(T_{ic} = 1|T_c = \pi)P(T_{jc} = 1|T_c = \pi)P(T_c = \pi) \\ &= \sum_{\Pi} \pi^2 f\left(\pi\right) \\ &= E[\pi^2] := \lambda = \eta^2 + \mu^2 \end{aligned}$$

  where the first equality follows from the chain rule of probability, $P(A_1, A_2, A_3) = P(A_1|A_2, A_3)P(A_2, A_3)$, and the second equality follows from the fact that randomization at the individual level is independent within a cluster i.e. T_{ic} is independent of T_{jc}, conditional on T_c.

- Defining λ in terms of μ and ρ yields $\lambda = \mu\left[\rho(1-\mu) + \mu\right]$

Using these expressions, compute the expectations:

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n} u_{ic}\right)^2\right] &= nE\left[u_{ic}^2\right] + n(n-1)E\left[u_{jc}u_{ic}\right] \\
&= n\left(\tau^2+\sigma^2\right) + n(n-1)\tau^2 \\
&= n^2\tau^2 + n\sigma^2 \\
E\left[\left(\sum_{i=1}^{n} u_{ic}\right)\left(\sum_{i=1}^{n} T_{ic}u_{ic}\right)\right] &= nE\left[u_{ic}^2 T_{ic}\right] + n(n-1)E\left[u_{jc}u_{ic}T_{ic}\right] \\
&= n\left(\tau^2+\sigma^2\right)\mu + n(n-1)\tau^2\mu \\
&= n^2\tau^2\mu + n\sigma^2\mu \\
E\left[\left(\sum_{i=1}^{n} T_{ic}u_{ic}\right)^2\right] &= nE\left[u_{ic}^2 T_{ic}^2\right] + n(n-1)E\left[u_{jc}u_{ic}T_{ic}T_{jc}\right] \\
&= n\left(\tau^2+\sigma^2\right)\mu + n(n-1)\tau^2\lambda
\end{aligned}
$$

Take the following expectations of matrices:

$$
\begin{aligned}
\frac{1}{n}E[X_c'X_c] &= \begin{bmatrix} 1 & \mu \\ \mu & \mu \end{bmatrix} \\
\frac{1}{n}E\left[X_c'u_cu_c'X_c\right] &= \begin{bmatrix} n\tau^2+\sigma^2 & \left(n\tau^2+\sigma^2\right)\mu \\ \left(n\tau^2+\sigma^2\right)\mu & \left(\tau^2+\sigma^2\right)\mu + (n-1)\tau^2\lambda \end{bmatrix}
\end{aligned}
$$

Plug these in to the relevant matrices necessary to calculate $SE\left(\widehat{\beta}_{OLS}\right)$. Note A is independent of whether one takes $n \to \infty$ or $C \to \infty$.

$$
\begin{aligned}
A &: = prob\lim \frac{1}{N}\sum_{c=1}^{C} X_c'X_c \\
&= \begin{bmatrix} 1 & \mu \\ \mu & \mu \end{bmatrix} \\
B &: = prob\lim \frac{1}{N}\sum_{c=1}^{C} X_c'u_cu_c'X_c \\
&= \begin{bmatrix} n\tau^2+\sigma^2 & \left(n\tau^2+\sigma^2\right)\mu \\ \left(n\tau^2+\sigma^2\right)\mu & \left(\tau^2+\sigma^2\right)\mu + (n-1)\tau^2\lambda_\pi \end{bmatrix}
\end{aligned}
$$

where the first equality is the definition and the second equality follows by plugging in the expressions for $\frac{1}{n}E[X_c'X_c]$ and $\frac{1}{n}E\left[X_c'u_cu_c'X_c\right]$, and the fact that all clusters are treated with identical probabilties ex-ante (i.e. these expressions are independent of c).

Now we need to invert A:

$$
A^{-1} = \frac{1}{\mu-\mu^2}\begin{bmatrix} \mu & -\mu \\ -\mu & 1 \end{bmatrix}
$$

Recall the expression for $SE\left(\widehat{\beta}_{OLS}\right)$

$$
\begin{aligned}
SE\left(\widehat{\beta}_{OLS}\right) \quad : \quad &= \frac{1}{N} * A^{-1}BA^{-1} \\
&= \frac{1}{nC} * \left(\frac{1}{\mu-\mu^2}\right)^2 \left[\begin{array}{cc} \mu & -\mu \\ -\mu & 1 \end{array}\right] \\
&\quad \times \left[\begin{array}{cc} n\tau^2+\sigma^2 & \left(n\tau^2+\sigma^2\right)\mu \\ \left(n\tau^2+\sigma^2\right)\mu & \left(\tau^2+\sigma^2\right)\mu+(n-1)\tau^2\lambda \end{array}\right] \times \left[\begin{array}{cc} \mu & -\mu \\ -\mu & 1 \end{array}\right] \\
&= \frac{1}{nC} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \\
&\quad \times \left[\begin{array}{cc} \mu\left(n\tau^2+\sigma^2\right)-\mu^2\left(n\tau^2+\sigma^2\right) & \left(n\tau^2+\sigma^2\right)\mu^2-\left(\tau^2+\sigma^2\right)\mu^2-(n-1)\tau^2\lambda\mu \\ 0 & \left(\tau^2+\sigma^2\right)\mu+(n-1)\tau^2\lambda-\left(n\tau^2+\sigma^2\right)\mu^2 \end{array}\right] \\
&\quad \times \left[\begin{array}{cc} \mu & -\mu \\ -\mu & 1 \end{array}\right] \\
&= \frac{1}{nC} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \\
&\quad \times \left[\begin{array}{cc} \mu\left(n\tau^2+\sigma^2\right)-\mu^2\left(n\tau^2+\sigma^2\right) & \left(n\tau^2+\sigma^2\right)\mu^2-\left(\tau^2+\sigma^2\right)\mu^2-(n-1)\tau^2\lambda\mu \\ 0 & \left(\tau^2+\sigma^2\right)\mu+(n-1)\tau^2\lambda-\left(n\tau^2+\sigma^2\right)\mu^2 \end{array}\right] \\
&\quad \times \left[\begin{array}{cc} \mu & -\mu \\ -\mu & 1 \end{array}\right] \\
&= \frac{1}{nC} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \\
&\quad \times \left[\begin{array}{cc} \mu^3\left(\tau^2+\sigma^2\right)-2\mu^3\left(n\tau^2+\sigma^2\right)+\mu^2\left(n\tau^2+\sigma^2\right)+(n-1)\mu^2\lambda\tau^2 & \mu^3\left(n\tau^2+\sigma^2\right)-\mu^2\left(\tau^2+\sigma^2\right) \\ \mu^3\left(n\tau^2+\sigma^2\right)-\mu^2\left(\tau^2+\sigma^2\right)-(n-1)\mu\lambda\tau^2 & -\mu^2\left(n\tau^2+\sigma^2\right)+\mu\left(\tau^2+\sigma^2\right) \end{array}\right.
\end{aligned}
$$

As we are interested particularly in $SE\left(\widehat{\beta}_1\right)$, we care about the bottom right value in the matrix:

$$
\begin{aligned}
SE\left(\widehat{\beta}_1\right) &= \frac{1}{nC} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \left(-\mu^2\left(n\tau^2+\sigma^2\right)+\mu\left(\tau^2+\sigma^2\right)+(n-1)\lambda\tau^2\right) \\
&= \frac{1}{nC} * \left(\frac{\left[\mu(1-n\mu)-\lambda(1-n)\right]\tau^2+\mu(1-\mu)\sigma^2}{\mu^2(1-\mu)^2}\right) \\
&= \frac{1}{nC} * \left[\left(\frac{\left[\mu(1-n\mu)-\lambda(1-n)\right]}{\mu^2(1-\mu)^2}\right)\tau^2+\left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right]
\end{aligned}
$$

Use $\lambda = \mu\left[\rho(1-\mu)+\mu\right]$ to express in terms of μ and ρ.

$$
SE\left(\widehat{\beta}_1\right) = \frac{1}{nC} * \left[\left(\frac{\left(1+\rho\left(n-1\right)\right)}{\mu\left(1-\mu\right)}\right)\tau^2+\left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right]
$$

8.2 Parameter values for different research designs:

Clustered Design

$$
\begin{aligned}
\Pi &= \{0,1\} \\
f(0) &= 1-P \\
f(1) &= P \\
\mu &= P \\
\rho &= 1
\end{aligned}
$$

Blocked Design

$$
\begin{aligned}
\Pi &= \{P\} \\
f(P) &= 1 \\
\mu &= P \\
\rho &= 0
\end{aligned}
$$

Randomized Saturation Design

$$
\begin{aligned}
\Pi &= \{\pi_1, ..., \pi_k\} \\
f(P) &= 1/k \\
\mu &= \frac{1}{k}\sum_{i=1}^{k}\pi_i \\
\eta^2 &= \frac{1}{k}\left(\sum_{i=1}^{k}\pi_i^2\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\pi_i\right)^2 \\
\rho &= \frac{\left(\sum_{i=1}^{k}\pi_i^2\right) - \frac{1}{k}\left(\sum_{i=1}^{k}\pi_i\right)^2}{\left(\sum_{i=1}^{k}\pi_i\right) - \frac{1}{k}\left(\sum_{i=1}^{k}\pi_i\right)^2}
\end{aligned}
$$

Partial Population Design

$$
\begin{aligned}
\Pi &= \{0, 0.5\} \\
f(\pi) &= 1/2 \\
\mu &= 1/4 \\
\eta^2 &= 1/16
\end{aligned}
$$

Now suppose 50% of clusters are treated at a randomized saturation and the remaining 50% of clusters are pure controls. Keep μ constant to avoid a power loss from changing the expected total

number of girls treated.

$$\begin{aligned} \Pi &= \{0, \pi_1, ..., \pi_k\} \\ f(0) &= 1/2 \\ f(\pi_i) &= p_i \\ \sum_k f(\pi_i) &= 1/2 \\ \mu &= \sum f(\pi_i)\pi_i \\ \eta^2 &= \sum f(\pi_i)\pi_i^2 - \mu^2 \end{aligned}$$

For a fixed μ, $SE\left(\widehat{\beta}_1\right)$ is minimized when η^2 is minimized. This is equivalent to minimizing

$$\begin{aligned} & \min_{f(\pi_i),\pi_i} \sum f(\pi_i)\pi_i^2 \\ s.t. \sum_k f(\pi_i) &= 1/2 \\ \sum_k f(\pi_i)\pi_i &= 1/4 \\ f(\pi_i) &\in [0, 1/2], \pi_i \in [0, 1] \end{aligned}$$

For simplicity, consider the case where each possible saturation is chosen with equal probability: $f(\pi_i) = 1/2k$. Then the problem simplifies to:

$$\begin{aligned} & \min_{\pi_i} \frac{1}{2k} \sum \pi_i^2 \\ s.t. \sum \pi_i &= k/2 \\ \pi_i &\in [0, 1] \end{aligned}$$

As the function π_i^2 is convex, and the constraint $\sum \pi_i = 2k$ implies that an increase in π_i must be offset by an accompanying decrease in $\sum_{j \neq i} \pi_j$, the minimum occurs at $\pi_i = \pi_j$ for all i, j. Plugging this into the equation $\sum \pi_i = 2k$ yields

$$\pi_i^* = 0.5$$

8.3 Proof: computing a closed form expression for $SE\left(\widehat{\beta}_{OLS}\right)$ when using only pure controls

Vary n_c. Note

$$N = \sum_{c=1}^{C} n_c$$

Using the underlying matrices of the model (specified in following section), we can write

$$\begin{aligned} X_c'X_c &= \begin{bmatrix} n_c & \sum_{i=1}^{n_c} T_{ic} \\ \sum_{i=1}^{n_c} T_{ic} & \sum_{i=1}^{n_c} T_{ic}^2 \end{bmatrix} \\ X_c'u_cu_c'X_c &= \begin{bmatrix} \left(\sum_{i=1}^{n_c} u_{ic}\right)^2 & \left(\sum_{i=1}^{n_c} u_{ic}\right)\left(\sum_{i=1}^{n_c} T_{ic}u_{ic}\right) \\ \left(\sum_{i=1}^{n_c} u_{ic}\right)\left(\sum_{i=1}^{n_c} T_{ic}u_{ic}\right) & \left(\sum_{i=1}^{n_c} T_{ic}u_{ic}\right)^2 \end{bmatrix} \end{aligned}$$

Using these expressions, compute the expectations:

$$\begin{aligned} E\left[\left(\sum_{i=1}^{n_c} u_{ic}\right)^2\right] &= n_c^2\tau^2 + n_c\sigma^2 \\ E\left[\left(\sum_{i=1}^{n_c} u_{ic}\right)\left(\sum_{i=1}^{n_c} T_{ic}u_{ic}\right)\right] &= n_c^2\tau^2\mu + n_c\sigma^2\mu \\ E\left[\left(\sum_{i=1}^{n_c} T_{ic}u_{ic}\right)^2\right] &= n_c\left(\tau^2+\sigma^2\right)\mu + n_c(n_c-1)\tau^2\lambda \end{aligned}$$

Take the following expectations of matrices:

$$
\begin{aligned}
\frac{1}{n_c}E[X_c'X_c] &= \begin{bmatrix} 1 & \mu \\ \mu & \mu \end{bmatrix} \\
E\left[X_c'u_cu_c'X_c\right] &= \begin{bmatrix} n_c^2\tau^2 + n_c\sigma^2 & n_c^2\tau^2\mu + n_c\sigma^2\mu \\ n_c^2\tau^2\mu + n_c\sigma^2\mu & n_c\left(\tau^2+\sigma^2\right)\mu + n_c(n_c-1)\tau^2\lambda \end{bmatrix} \\
\sum_{c=1}^{C} E\left[X_c'u_cu_c'X_c\right] &= \begin{bmatrix} \tau^2\sum_{c=1}^{C} n_c^2 + \sigma^2 N & \tau^2\mu\sum_{c=1}^{C} n_c^2 + \sigma^2\mu N \\ \tau^2\mu\sum_{c=1}^{C} n_c^2 + \sigma^2\mu N & N\left(\tau^2+\sigma^2\right)\mu + \tau^2\lambda\sum_{c=1}^{C} n_c^2 - N\tau^2\lambda \end{bmatrix} \\
\frac{1}{N}\sum_{c=1}^{C} E\left[X_c'u_cu_c'X_c\right] &= \begin{bmatrix} \tau^2\phi + \sigma^2 & \tau^2\mu\phi + \sigma^2\mu \\ \tau^2\mu\phi + \sigma^2\mu & \left(\tau^2+\sigma^2\right)\mu + \tau^2\lambda\phi - \tau^2\lambda \end{bmatrix}
\end{aligned}
$$

Where

$$\phi = \frac{\sum_{c=1}^{C} n_c^2}{N}$$

is the average of n_c^2. Note ϕ is minimized at $n_c = n$ for all c, and is equal to $\phi = n$.

Plug these in to the relevant matrices necessary to calculate $SE\left(\widehat{\beta}_{OLS}\right)$. Note A is independent of whether one takes $n_c \to \infty$ or $C \to \infty$.

$$
\begin{aligned}
A &: = prob\lim \frac{1}{N}\sum_{c=1}^{C} X_c'X_c \\
&= \begin{bmatrix} 1 & \mu \\ \mu & \mu \end{bmatrix} \\
B &: = prob\lim \frac{1}{N}\sum_{c=1}^{C} X_c'u_cu_c'X_c \\
&= \begin{bmatrix} \phi\tau^2 + \sigma^2 & \left(\phi\tau^2+\sigma^2\right)\mu \\ \left(\phi\tau^2+\sigma^2\right)\mu & \left(\tau^2+\sigma^2\right)\mu + (\phi-1)\tau^2\lambda \end{bmatrix}
\end{aligned}
$$

where the first equality is the definition and the second equality follows by plugging in the expressions for $E[X_c'X_c]$ and $E\left[X_c'u_cu_c'X_c\right]$, and the fact that all clusters are treated with identical probabilties ex-ante (i.e. these expressions are independent of c).

Now we need to invert A: good thing it is a $2x2$ matrix!

$$A^{-1} = \frac{1}{\mu - \mu^2} \begin{bmatrix} \mu & -\mu \\ -\mu & 1 \end{bmatrix}$$

Recall the expression for $SE\left(\widehat{\beta}_{OLS}\right)$

$$\begin{aligned}
SE\left(\widehat{\beta}_{OLS}\right) \quad : \quad &= \frac{1}{N} * A^{-1} B A^{-1} \\
&= \frac{1}{N} * \left(\frac{1}{\mu - \mu^2}\right)^2 \begin{bmatrix} \mu & -\mu \\ -\mu & 1 \end{bmatrix} \\
&\quad \times \begin{bmatrix} \phi\tau^2 + \sigma^2 & \left(\phi\tau^2 + \sigma^2\right)\mu \\ \left(\phi\tau^2 + \sigma^2\right)\mu & \left(\tau^2 + \sigma^2\right)\mu + (\phi - 1)\tau^2\lambda \end{bmatrix} \times \begin{bmatrix} \mu & -\mu \\ -\mu & 1 \end{bmatrix} \\
&= \frac{1}{N} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \\
&\quad \times \begin{bmatrix} \mu\left(\phi\tau^2 + \sigma^2\right) - \mu^2\left(\phi\tau^2 + \sigma^2\right) & \left(\phi\tau^2 + \sigma^2\right)\mu^2 - \left(\tau^2 + \sigma^2\right)\mu^2 - (\phi - 1)\tau^2\lambda\mu \\ 0 & \left(\tau^2 + \sigma^2\right)\mu + (\phi - 1)\tau^2\lambda - \left(\phi\tau^2 + \sigma^2\right)\mu^2 \end{bmatrix} \\
&\quad \times \begin{bmatrix} \mu & -\mu \\ -\mu & 1 \end{bmatrix} \\
&= \frac{1}{N} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \\
&\quad \times \begin{bmatrix} \mu\left(\phi\tau^2 + \sigma^2\right) - \mu^2\left(\phi\tau^2 + \sigma^2\right) & \left(\phi\tau^2 + \sigma^2\right)\mu^2 - \left(\tau^2 + \sigma^2\right)\mu^2 - (\phi - 1)\tau^2\lambda\mu \\ 0 & \left(\tau^2 + \sigma^2\right)\mu + (\phi - 1)\tau^2\lambda - \left(\phi\tau^2 + \sigma^2\right)\mu^2 \end{bmatrix} \\
&\quad \times \begin{bmatrix} \mu & -\mu \\ -\mu & 1 \end{bmatrix} \\
&= \frac{1}{N} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \\
&\quad \times \begin{bmatrix} \mu^3\left(\tau^2 + \sigma^2\right) - 2\mu^3\left(\phi\tau^2 + \sigma^2\right) + \mu^2\left(\phi\tau^2 + \sigma^2\right) + (\phi - 1)\mu^2\lambda\tau^2 & \mu^3\left(\phi\tau^2 + \sigma^2\right) - \mu^2\left(\tau^2 + \right. \\ \mu^3\left(\phi\tau^2 + \sigma^2\right) - \mu^2\left(\tau^2 + \sigma^2\right) - (\phi - 1)\mu\lambda\tau^2 & -\mu^2\left(\phi\tau^2 + \sigma^2\right) + \mu\left(\tau^2 + \right.
\end{aligned}$$

As we are interested particularly in $SE\left(\widehat{\beta}_1\right)$, we care about the bottom right value in the matrix:

$$
\begin{aligned}
SE\left(\widehat{\beta}_1\right) &= \frac{1}{N} * \left(\frac{1}{\mu(1-\mu)}\right)^2 \left(-\mu^2\left(\phi\tau^2+\sigma^2\right)+\mu\left(\tau^2+\sigma^2\right)+(\phi-1)\lambda\tau^2\right) \\
&= \frac{1}{N} * \left(\frac{\left[\mu(1-\phi\mu)-\lambda(1-\phi)\right]\tau^2+\mu(1-\mu)\sigma^2}{\mu^2(1-\mu)^2}\right) \\
&= \frac{1}{N} * \left[\left(\frac{\left[\mu(1-\phi\mu)-\lambda(1-\phi)\right]}{\mu^2(1-\mu)^2}\right)\tau^2+\left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right]
\end{aligned}
$$

Use $\lambda = \mu\left[\rho(1-\mu)+\mu\right]$ to express in terms of μ and ρ.

$$
SE\left(\widehat{\beta}_1\right) = \frac{1}{N} * \left[\left(\frac{\left(1+\rho\left(\phi-1\right)\right)}{\mu\left(1-\mu\right)}\right)\tau^2+\left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right]
$$

Look at

$$
N = n(1-\mu)C + \sum_{c=1}^{\mu C} n_c
$$

to see if $\mu > 1/2$ is now optimal to offset some of the power loss among treated girls from varying n_c.

TABLE 1.

Baseline Covariates:

Variable	Obs	Mean	Std. Dev.	Min	Max
Household Size	2651	6.436	2.207	1	15
Household Asset Index	2651	0.813	2.611	-3.697147	6.827225
Mobie Phone Ownership	2651	0.619	0.486	0	1
Age at Baseline	2653	15.214	1.900	13	22
Mother Alive	2653	0.836	0.371	0	1
Father Alive	2648	0.715	0.451	0	1
Never had Sex at baseline	2653	0.801	0.399	0	1
Ever Pregnant at baseline	2652	0.025	0.155	0	1
Spending on girl at baseline	2653	921.886	1510.221	0	27880
Spending by girl at baseline	2653	16.735	120.927	0	4620

Endline Outcomes:

Variable	Obs	Mean	Std. Dev.	Min	Max
Terms in school 2008	2582	2.671	0.849	0	3
English test score	2615	0.008	1.008	-2.270203	2.287322
Cognitive test score	2615	0.077	1.026	-1.566874	2.631495
Ever Married at endline	2652	0.164	0.370	0	1
Ever Pregnant at endline	2653	0.236	0.425	0	1
HIV Positive at endline	2575	0.025	0.155	0	1
Spending on girl at endline	2652	1271.919	1729.605	0	23600
Spending by girl at endline	2652	23.644	167.577	0	7100

TABLE 2.

Balance Test

	Dependent Variable at Baseline:									
	Household Size	Asset Index	Mobile Phone Ownership	Age	Mother Alive	Father Alive	Never had Sex	Ever Pregnant	Expenditures on girls' own consumption	Money spent by girl on other girls' consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CCT	0.031	0.624	0.001	-0.343	-0.038	-0.004	0.009	0.004	1.082	-0.022
	(0.216)	(0.478)	(0.059)	(0.150)**	(0.027)	(0.038)	(0.031)	(0.008)	(1.226)	(0.055)
UCT	0.275	0.672	-0.001	0.130	-0.005	0.048	-0.016	0.006	0.144	-0.062
	(0.186)	(0.390)*	(0.067)	(0.122)	(0.026)	(0.034)	(0.036)	(0.009)	(0.656)	(0.047)
Within CCT EA Control	-0.211	-0.167	-0.024	0.090	-0.035	0.036	-0.003	0.006	1.820	0.044
	(0.205)	(0.508)	(0.073)	(0.154)	(0.033)	(0.036)	(0.041)	(0.018)	(1.166)	(0.106)
Within UCT EA Control	0.416	0.086	-0.048	0.295	0.067	-0.111	0.103	-0.056	-0.157	0.012
	(0.528)	(1.337)	(0.206)	(0.596)	(0.110)	(0.163)	(0.071)	(0.054)	(1.540)	(0.067)
EA CCT Treatment Saturation	0.207	1.060	0.106	0.093	0.146	0.036	0.048	-0.005	-1.497	-0.141
	(0.452)	(1.223)	(0.167)	(0.388)	(0.073)**	(0.089)	(0.104)	(0.040)	(2.592)	(0.241)
EA UCT Treatment Saturation	-1.182	0.943	0.249	-0.664	-0.118	0.249	-0.176	0.135	-1.697	-0.161
	(0.990)	(2.669)	(0.445)	(1.142)	(0.200)	(0.261)	(0.172)	(0.156)	(3.154)	(0.108)
Observations	2,651	2,651	2,651	2,653	2,653	2,648	2,653	2,652	2,653	2,653
R-squared	0	0.01	0	0.01	0	0	0	0	0	0

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation of one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

TABLE 3.

Basic Spillover Analysis

	Dependent Variable:															
	Terms Enrolled		English Test Scores		Cognitive Test Scores		Ever Married		Ever Pregnant		HIV Positive		Expenditures on girls' own consumption		Money spent by girl on other girls' consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Treatment	0.107	0.252	0.051	0.099	0.151	0.234	-0.027	-0.068	-0.007	-0.017	0.000	-0.003	3.594	3.570	0.141	0.378
	(0.039)***	(0.085)***	(0.044)	(0.129)	(0.048)***	(0.143)	(0.017)	(0.054)	(0.022)	(0.059)	(0.009)	(0.021)	(0.838)***	(1.799)**	(0.070)**	(0.201)*
Within-EA Control	-0.010	0.000	-0.016	-0.119	0.093	-0.009	0.005	0.009	-0.003	-0.002	0.009	0.014	0.968	1.034	0.055	-0.079
	(0.041)	(0.073)	(0.076)	(0.096)	(0.075)	(0.091)	(0.020)	(0.031)	(0.020)	(0.029)	(0.011)	(0.014)	(0.872)	(1.276)	(0.047)	(0.045)*
EA saturation,		-0.187		-0.061		-0.107		0.053		0.014		0.003		0.031		-0.306
for Treatments		(0.105)*		(0.152)		(0.180)		(0.062)		(0.076)		(0.028)		(2.472)		(0.200)
EA saturation,		-0.035		0.345		0.342		-0.012		-0.002		-0.015		-0.221		0.443
Within-EA Control		(0.173)		(0.315)		(0.295)		(0.063)		(0.071)		(0.031)		(2.903)		(0.225)*
Observations	2,579	2,579	2,612	2,612	2,612	2,612	2,649	2,649	2,650	2,650	2,572	2,572	2,650	2,650	2,650	2,650
R-squared	0.09	0.1	0.41	0.41	0.18	0.18	0.14	0.14	0.19	0.19	0.03	0.03	0.1	0.1	0.03	0.03
Estimates of:																
E(ITT)	0.107		0.051		0.151		-0.027		-0.007		0.000		3.594		0.141	
E(ToT)	0.146		0.073		0.170		-0.038		-0.008		-0.003		4.469		0.170	
TUT	0.336		0.132		0.312		-0.091		-0.023		-0.004		4.760		0.504	
E(ASNT)	-0.010		-0.016		0.093		0.005		-0.003		0.009		0.968		0.055	
E(AST)	-0.19		-0.059		-0.142		0.053		0.014		0.001		-0.291		-0.334	

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation of one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence. Estimates at the bottom of the table are based on an aggregate round 3 compliance rate of 75% between CCT and UCT.

TABLE 4.

Basic Spillover Analysis, Split

	Dependent Variable:															
	Terms Enrolled		English Test Scores		Cognitive Test Scores		Ever Married		Ever Pregnant		HIV Positive		Expenditures on girls' own consumption		Money spent by girl on other girls' consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
CCT	0.134	0.276	0.084	0.296	0.166	0.264	-0.009	-0.018	0.016	0.060	0.009	0.004	4.474	4.476	0.150	0.460
	(0.045)***	(0.121)**	(0.048)*	(0.129)**	(0.052)***	(0.131)**	(0.019)	(0.073)	(0.027)	(0.079)	(0.012)	(0.030)	(0.963)***	(2.218)**	(0.099)	(0.275)*
UCT	0.050	0.214	-0.021	-0.202	0.117	0.189	-0.064	-0.140	-0.054	-0.130	-0.019	-0.013	1.765	2.441	0.122	0.254
	(0.050)	(0.100)**	(0.073)	(0.201)	(0.081)	(0.292)	(0.025)**	(0.047)***	(0.023)**	(0.056)**	(0.008)**	(0.014)	(1.024)*	(1.780)	(0.078)	(0.316)
Within CCT EA Control	0.017	0.003	0.000	-0.100	0.192	0.037	0.007	0.007	0.000	-0.007	0.000	0.008	1.657	1.519	0.071	-0.091
	(0.047)	(0.077)	(0.082)	(0.092)	(0.072)***	(0.079)	(0.023)	(0.032)	(0.024)	(0.031)	(0.009)	(0.014)	(1.019)	(1.342)	(0.060)	(0.042)**
Within UCT EA Control	-0.080	0.019	-0.063	-0.352	-0.151	-0.330	0.002	0.047	-0.009	0.083	0.032	0.067	-0.739	-3.454	0.014	0.165
	(0.066)	(0.205)	(0.148)	(0.370)	(0.138)	(0.397)	(0.032)	(0.087)	(0.030)	(0.079)	(0.026)	(0.068)	(0.927)	(2.054)*	(0.063)	(0.157)
EA saturation,		-0.181		-0.270		-0.124		0.011		-0.056		0.005		0.006		-0.394
CCT Treatment		(0.154)		(0.153)*		(0.187)		(0.082)		(0.104)		(0.041)		(2.920)		(0.255)
EA saturation,		-0.217		0.244		-0.097		0.102		0.102		-0.008		-0.908		-0.177
UCT treatment		(0.134)		(0.256)		(0.346)		(0.068)		(0.065)		(0.014)		(2.688)		(0.351)
EA saturation,		0.054		0.394		0.610		0.001		0.029		-0.029		0.539		0.627
Within-CCT Control		(0.169)		(0.415)		(0.272)**		(0.069)		(0.085)		(0.032)		(3.563)		(0.283)**
EA saturation,		-0.242		0.708		0.439		-0.111		-0.224		-0.086		6.635		-0.368
Within-UCT Control		(0.529)		(0.622)		(0.875)		(0.175)		(0.153)		(0.121)		(3.716)*		(0.254)
Observations	2,579	2,579	2,612	2,612	2,612	2,612	2,649	2,649	2,650	2,650	2,572	2,572	2,650	2,650	2,650	2,650
R-squared	0.1	0.1	0.41	0.41	0.18	0.19	0.14	0.14	0.19	0.19	0.03	0.03	0.11	0.11	0.03	0.03
Estimates for CCT::																
E(ITT)	0.134		0.084		0.166		-0.009		0.016		0.009		4.474		0.150	
E(ToT)	0.210		0.139		0.149		-0.019		0.026		0.015		6.313		0.202	
TUT	0.456		0.489		0.436		-0.030		0.099		0.007		7.398		0.760	
E(ASNT)	0.017		0.000		0.192		0.007		0.000		0.000		1.657		0.071	
E(AST)	-0.246		-0.350		-0.287		0.010		-0.073		0.008		-1.085		-0.559	
Estimates for UCT::																
E(ITT)	0.050		-0.021		0.117		-0.064		-0.054		-0.019		1.765		0.122	
E(ToT)	0.051		-0.021		0.120		-0.065		-0.054		-0.020		1.790		0.123	
TUT	0.279		0.299		0.267		-0.018		0.061		0.004		4.521		0.465	
E(ASNT)	-0.080		-0.063		-0.151		0.002		-0.009		0.032		-0.739		0.014	
E(AST)	-0.227		-0.320		-0.147		-0.046		-0.115		-0.024		-2.731		-0.342	

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation of one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

TABLE 5.

Granular Spillover Analysis

	Dependent Variable:							
	Terms Enrolled	English Test Scores	Cognitive Test Scores	Ever Married	Ever Pregnant	HIV Positive	Expenditures on girls' own consumption	Money spent by girl on other girls' consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment 33%	0.185	0.084	0.196	-0.033	0.015	-0.014	2.627	0.140
	(0.061)***	(0.106)	(0.111)*	(0.044)	(0.042)	(0.009)	(0.946)***	(0.111)
Treatment 66%	0.132	0.052	0.166	-0.051	-0.038	0.013	4.632	0.324
	(0.055)**	(0.060)	(0.049)***	(0.021)**	(0.027)	(0.015)	(1.295)***	(0.181)*
Treatment 100%	0.064	0.039	0.127	-0.009	0.006	-0.004	3.274	0.025
	(0.051)	(0.055)	(0.072)*	(0.019)	(0.031)	(0.013)	(1.256)***	(0.045)
Spillover 0%	0.002	-0.089	0.099	0.011	0.008	0.014	1.290	-0.040
	(0.084)	(0.093)	(0.077)	(0.035)	(0.032)	(0.015)	(1.414)	(0.027)
Spillover 33%	-0.014	-0.042	-0.034	0.003	-0.016	0.008	0.634	0.020
	(0.045)	(0.101)	(0.097)	(0.028)	(0.026)	(0.018)	(1.219)	(0.043)
Spillover 66%	-0.020	0.156	0.382	0.003	0.011	0.005	1.285	0.276
	(0.080)	(0.215)	(0.182)**	(0.026)	(0.039)	(0.016)	(1.449)	(0.163)*
Observations	2,579	2,612	2,612	2,649	2,650	2,572	2,650	2,650

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation of one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

TABLE 6.

Extended Spillover Analysis

	Dependent Variable:							
	Terms Enrolled	English Test Scores	Cognitive Test Scores	Ever Married	Ever Pregnant	HIV Positive	Expenditures on girls' own consumption	Money spent by girl on other girls' consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CCT 33%	0.180	0.243	0.255	0.021	0.089	-0.006	3.613	0.104
	(0.085)**	(0.097)**	(0.089)***	(0.057)	(0.049)*	(0.014)	(1.198)***	(0.061)*
CCT 66%	0.185	0.087	0.157	-0.042	-0.018	0.019	5.210	0.393
	(0.054)***	(0.066)	(0.051)***	(0.024)*	(0.032)	(0.019)	(1.480)***	(0.243)
CCT 100%	0.084	0.038	0.151	0.003	0.018	0.006	4.243	0.004
	(0.063)	(0.059)	(0.086)*	(0.021)	(0.042)	(0.019)	(1.408)***	(0.048)
UCT 33%	0.192	-0.122	0.118	-0.104	-0.080	-0.025	1.307	0.186
	(0.059)***	(0.146)	(0.220)	(0.030)***	(0.038)**	(0.008)***	(0.922)	(0.242)
UCT 66%	-0.003	-0.036	0.191	-0.076	-0.089	-0.002	3.226	0.150
	(0.084)	(0.088)	(0.088)**	(0.034)**	(0.039)**	(0.016)	(1.835)*	(0.148)
UCT 100%	0.020	0.041	0.072	-0.037	-0.020	-0.026	1.111	0.073
	(0.069)	(0.105)	(0.109)	(0.036)	(0.026)	(0.007)***	(1.489)	(0.068)
Spillover CCT 0%	0.002	-0.088	0.100	0.011	0.008	0.014	1.301	-0.041
	(0.084)	(0.093)	(0.077)	(0.035)	(0.032)	(0.015)	(1.417)	(0.026)
Spillover CCT 33%	0.026	0.020	0.097	-0.004	-0.036	-0.017	2.187	0.005
	(0.055)	(0.075)	(0.079)	(0.038)	(0.033)	(0.009)*	(1.580)	(0.034)
Spillover CCT 66%	0.035	0.173	0.577	0.017	0.046	0.003	1.454	0.436
	(0.072)	(0.309)	(0.163)***	(0.029)	(0.053)	(0.018)	(2.005)	(0.212)**
Spillover UCT 33%	-0.061	-0.116	-0.185	0.010	0.009	0.038	-1.254	0.041
	(0.063)	(0.180)	(0.158)	(0.038)	(0.036)	(0.032)	(1.025)	(0.077)
Spillover UCT 66%	-0.141	0.120	-0.036	-0.027	-0.067	0.011	0.983	-0.075
	(0.166)	(0.115)	(0.249)	(0.047)	(0.040)*	(0.027)	(0.981)	(0.038)*
Observations	2,579	2,612	2,612	2,649	2,650	2,572	2,650	2,650

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation of one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

TABLE 7.

Analysis of Household Networks, Treatments Pooled.

	Terms Enrolled	English Test Scores	Cognitive Test Scores	Ever Married	Ever Pregnant	HIV Positive	Money spent by beneficiary on own consumption	Money spent on beneficiary's consumption by others
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.09**	0.05	0.12*	-0.02	-0.00	-0.01	0.95	-0.03
	(0.043)	(0.070)	(0.064)	(0.020)	(0.023)	(0.008)	(0.796)	(0.053)
Within-Village Control	-0.01	0.02	0.12	0.01	0.00	0.01	0.32	-0.05
	(0.051)	(0.103)	(0.076)	(0.023)	(0.023)	(0.013)	(0.781)	(0.046)
Number of Treated HH members	0.12**	-0.01	0.10	-0.03	-0.03	0.03	2.26	0.34
For Treated Girls	(0.062)	(0.175)	(0.096)	(0.020)	(0.027)	(0.027)	(1.735)	(0.230)
Number of Treated HH members	0.10	-0.18	-0.12	-0.02	-0.03	-0.04***	0.83	-0.08
For Untreated Girls	(0.102)	(0.136)	(0.166)	(0.041)	(0.042)	(0.013)	(1.467)	(0.056)
Number of HH members in Sampling Frame	0.00	0.14***	0.05	-0.01	-0.01	0.01*	0.18	0.03
	(0.029)	(0.040)	(0.047)	(0.011)	(0.014)	(0.008)	(0.496)	(0.060)
Number of dropout HH members	-0.06	0.27***	0.23**	-0.05***	-0.07**	-0.00	4.30***	0.07
	(0.059)	(0.067)	(0.099)	(0.019)	(0.027)	(0.016)	(1.427)	(0.151)
Average Sampling Weight in Household	-0.08	-0.70***	-0.64***	0.22***	0.23***	-0.02	-5.35***	0.02
	(0.081)	(0.120)	(0.091)	(0.027)	(0.034)	(0.015)	(1.381)	(0.062)
Constant	2.69***	0.36***	0.39***	0.03	0.09***	0.02*	11.46***	0.11**
	(0.076)	(0.099)	(0.080)	(0.026)	(0.032)	(0.014)	(1.352)	(0.044)
Observations	2,658	2,618	2,618	2,650	2,651	2,581	2,655	2,655
R-squared	0.009	0.081	0.057	0.041	0.034	0.014	0.050	0.012

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

TABLE 8.

Analysis of Household Networks, Treatments Split.

	Terms Enrolled	English Test Scores	Cognitive Test Scores	Ever Married	Ever Pregnant	HIV Positive	Money spent by beneficiary on own consumption	Money spent on beneficiary's consumption by others
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CCT Treatment	0.12***	0.01	0.08	0.01	0.03	-0.01	0.75	0.02
	(0.045)	(0.076)	(0.065)	(0.023)	(0.027)	(0.011)	(1.035)	(0.068)
UCT Treatment	0.02	0.12	0.18	-0.07***	-0.06**	-0.02**	1.31	-0.11***
	(0.063)	(0.104)	(0.117)	(0.022)	(0.029)	(0.006)	(0.818)	(0.040)
Within-Village Control CCT	0.02	-0.00	0.18**	0.01	0.00	0.00	0.59	-0.04
	(0.059)	(0.133)	(0.081)	(0.027)	(0.027)	(0.010)	(0.926)	(0.052)
Within-Village Control UCT	-0.07	0.08	-0.06	-0.00	-0.01	0.05	-0.49	-0.08*
	(0.066)	(0.116)	(0.124)	(0.034)	(0.032)	(0.032)	(0.737)	(0.044)
Number of CCT-Treated HH Members	0.09	-0.00	0.12	-0.06**	-0.05**	0.03	2.35	0.41
for Treated Girls	(0.070)	(0.219)	(0.109)	(0.023)	(0.027)	(0.033)	(2.092)	(0.284)
Number of UCT-Treated HH Members	0.22***	-0.00	0.08	0.02	0.02	-0.00	2.19	-0.02
for Treated Girls	(0.078)	(0.152)	(0.146)	(0.037)	(0.054)	(0.013)	(2.611)	(0.062)
Number of CCT-Treated HH Members	0.05	-0.12	0.00	-0.01	-0.02	-0.03***	2.18	-0.12***
for Untreated Girls	(0.151)	(0.222)	(0.239)	(0.066)	(0.070)	(0.011)	(1.812)	(0.042)
Number of UCT-Treated HH Members	0.18*	-0.26*	-0.17	-0.02	-0.05	-0.07***	-0.06	-0.03
for Untreated Girls	(0.103)	(0.138)	(0.181)	(0.047)	(0.035)	(0.026)	(1.870)	(0.066)
Number of HH members in Sampling Frame	0.00	0.14***	0.05	-0.01	-0.01	0.01*	0.17	0.03
	(0.029)	(0.040)	(0.048)	(0.011)	(0.014)	(0.007)	(0.501)	(0.062)
Number of dropout HH members	-0.06	0.27***	0.24**	-0.05***	-0.07**	-0.00	4.34***	0.07
	(0.059)	(0.068)	(0.103)	(0.019)	(0.027)	(0.015)	(1.452)	(0.151)
Average Sampling Weight in Household	-0.08	-0.70***	-0.64***	0.22***	0.23***	-0.01	-5.36***	0.05
	(0.081)	(0.123)	(0.089)	(0.027)	(0.033)	(0.014)	(1.358)	(0.060)
Constant	2.69***	0.36***	0.40***	0.03	0.09***	0.02	11.47***	0.09**
	(0.076)	(0.100)	(0.080)	(0.026)	(0.031)	(0.014)	(1.345)	(0.045)
Observations	2,658	2,618	2,618	2,650	2,651	2,581	2,655	2,655
R-squared	0.010	0.082	0.061	0.044	0.036	0.020	0.051	0.016

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

TABLE 9.

Analysis of Social Networks, Treatments Pooled.

	Terms Enrolled	English Test Scores	Cognitive Test Scores	Ever Married	Ever Pregnant	HIV Positive	Money spent by beneficiary on own consumption	Money spent on beneficiary's consumption by others
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CCT	0.18***	0.12	0.21*	-0.03	-0.02	0.01	1.99	0.20*
	(0.057)	(0.116)	(0.109)	(0.029)	(0.035)	(0.016)	(1.281)	(0.114)
UCT	0.10	0.13	0.21*	-0.08***	-0.07***	-0.02**	1.40	-0.12***
	(0.063)	(0.129)	(0.107)	(0.025)	(0.026)	(0.008)	(1.135)	(0.046)
Within-Village Control CCT	0.02	0.01	0.21***	0.01	-0.01	0.00	1.10	-0.04
	(0.058)	(0.143)	(0.078)	(0.027)	(0.025)	(0.010)	(1.089)	(0.048)
Within-Village Control UCT	-0.06	0.04	-0.08	-0.02	-0.05	0.04	-0.28	-0.09*
	(0.076)	(0.099)	(0.091)	(0.034)	(0.041)	(0.029)	(0.953)	(0.048)
Number of Treated Friends for Treatment Girls	-0.02	-0.12*	-0.11	0.01	0.02	-0.01	-0.37	-0.06
	(0.050)	(0.069)	(0.078)	(0.019)	(0.023)	(0.007)	(0.683)	(0.050)
Number of Treated Friends for Untreated Girls	0.11*	-0.03	-0.03	-0.02	0.03	-0.02**	-1.03	-0.07
	(0.067)	(0.080)	(0.074)	(0.028)	(0.037)	(0.010)	(0.712)	(0.047)
Number of friends who are dropouts	-0.16***	-0.10**	-0.14***	0.08***	0.12***	0.02	-0.58	0.08
	(0.047)	(0.047)	(0.036)	(0.021)	(0.022)	(0.013)	(0.428)	(0.070)
Number of friends in same cluster	0.01	-0.10***	-0.09***	0.00	-0.01	-0.00**	-1.09***	-0.03**
	(0.015)	(0.021)	(0.018)	(0.006)	(0.006)	(0.002)	(0.190)	(0.014)
1 Matched Friend	-0.01	-0.11*	-0.14***	0.04*	0.03	0.00	-0.47	0.01
	(0.044)	(0.056)	(0.049)	(0.019)	(0.022)	(0.008)	(0.738)	(0.054)
2 Matched Friends	-0.00	-0.01	0.02	0.05*	0.04	-0.01*	-0.34	-0.06
	(0.068)	(0.080)	(0.064)	(0.031)	(0.030)	(0.006)	(0.931)	(0.039)
3 Matched Friends	-0.02	0.00	-0.07	0.08**	0.07*	-0.01	0.64	0.18
	(0.097)	(0.117)	(0.084)	(0.038)	(0.041)	(0.009)	(1.088)	(0.188)
4 Matched Friends	-0.10	0.03	-0.00	0.09	0.06	0.04	-0.89	0.05
	(0.197)	(0.172)	(0.154)	(0.076)	(0.088)	(0.038)	(1.466)	(0.100)
5 Matched Friends	0.23*	-0.05	0.50***	-0.15***	-0.24***	0.01	-2.76	0.01
	(0.126)	(0.342)	(0.178)	(0.051)	(0.058)	(0.014)	(2.008)	(0.123)
Constant	2.63***	0.19***	0.21***	0.14***	0.22***	0.03***	10.50***	0.20***
	(0.047)	(0.066)	(0.052)	(0.019)	(0.019)	(0.007)	(0.843)	(0.039)
Observations	2,660	2,620	2,620	2,652	2,653	2,583	2,657	2,657
R-squared	0.014	0.042	0.049	0.021	0.022	0.013	0.030	0.010

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

TABLE 10.

Analysis of Social Networks, Treatments Split

	Terms Enrolled	English Test Scores	Cognitive Test Scores	Ever Married	Ever Pregnant	HIV Positive	Money spent by beneficiary on own consumption	Money spent on beneficiary's consumption by others
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CCT	0.26***	0.07	0.21	-0.04	0.02	0.01	5.05*	-0.10
	(0.097)	(0.238)	(0.188)	(0.055)	(0.062)	(0.021)	(2.889)	(0.124)
UCT	0.18*	0.08	0.21	-0.09*	-0.04	-0.02	4.31	-0.40***
	(0.096)	(0.200)	(0.213)	(0.050)	(0.055)	(0.021)	(2.812)	(0.149)
Within-Village Control CCT	0.04	-0.04	0.10	0.01	-0.01	0.01	-0.52	-0.05
	(0.085)	(0.148)	(0.086)	(0.038)	(0.032)	(0.014)	(0.896)	(0.054)
Within-Village Control UCT	-0.03	-0.04	-0.26*	-0.02	-0.05	0.05	-2.83**	-0.10
	(0.106)	(0.232)	(0.136)	(0.048)	(0.052)	(0.033)	(1.149)	(0.082)
Number of Treated Friends	-0.01	-0.13*	-0.12	0.01	0.02	-0.01	-0.14	-0.10*
for Treatment Girls	(0.053)	(0.074)	(0.089)	(0.019)	(0.025)	(0.008)	(0.631)	(0.054)
Number of Treated Friends	0.12*	-0.05	-0.06	-0.02	0.03	-0.02*	-1.39**	-0.08
for Untreated Girls	(0.066)	(0.089)	(0.075)	(0.028)	(0.037)	(0.010)	(0.702)	(0.051)
EA saturation,	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.04	0.00*
for Treatments	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)	(0.000)	(0.032)	(0.002)
EA saturation,	-0.00	0.00	0.00*	0.00	-0.00	-0.00	0.07***	0.00
Within-EA Control	(0.002)	(0.005)	(0.003)	(0.001)	(0.001)	(0.000)	(0.019)	(0.002)
Number of friends who are dropouts	-0.16***	-0.10**	-0.14***	0.08***	0.12***	0.02	-0.49	0.08
	(0.047)	(0.047)	(0.037)	(0.021)	(0.022)	(0.013)	(0.417)	(0.071)
Number of friends in same cluster	0.01	-0.10***	-0.09***	0.00	-0.01	-0.00**	-1.08***	-0.03**
	(0.015)	(0.020)	(0.018)	(0.006)	(0.006)	(0.002)	(0.187)	(0.014)
1 Matched Friend	-0.01	-0.11*	-0.13***	0.04*	0.03	0.00	-0.43	0.01
	(0.044)	(0.054)	(0.049)	(0.019)	(0.022)	(0.008)	(0.749)	(0.055)
2 Matched Friends	-0.01	0.00	0.03	0.05*	0.04	-0.01**	-0.33	-0.04
	(0.068)	(0.077)	(0.064)	(0.031)	(0.030)	(0.007)	(0.947)	(0.039)
3 Matched Friends	-0.03	0.01	-0.06	0.08**	0.07*	-0.01	0.65	0.20
	(0.098)	(0.115)	(0.087)	(0.038)	(0.042)	(0.010)	(1.117)	(0.189)
4 Matched Friends	-0.11	0.05	0.02	0.09	0.06	0.04	-0.74	0.07
	(0.198)	(0.171)	(0.154)	(0.076)	(0.088)	(0.038)	(1.520)	(0.102)
5 Matched Friends	0.18	0.01	0.58***	-0.15***	-0.25***	0.00	-2.43	0.11
	(0.137)	(0.335)	(0.183)	(0.054)	(0.066)	(0.018)	(1.715)	(0.147)
Constant	2.64***	0.19***	0.21***	0.14***	0.22***	0.03***	10.48***	0.20***
	(0.048)	(0.065)	(0.051)	(0.019)	(0.019)	(0.007)	(0.851)	(0.037)
Observations	2,660	2,620	2,620	2,652	2,653	2,583	2,657	2,657
R-squared	0.014	0.043	0.051	0.021	0.022	0.013	0.035	0.012

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

TABLE 11. Change in Social Networks between Round 1 and Round 2.

		Schoolgirls		
		Control	Treatment	p-value
Average changes between R1 and R2	# friends listed	0.14	0.10	0.5088
	# friends in sampling frame	-0.02	-0.09	0.4141
	# friends matched	-0.12	0.03	0.0010
	# friends treated dropouts	0.01	0.03	0.0017
	# friends treated schoolgirls	-0.01	0.03	0.0240
	# friends treated	0.00	0.06	0.0112
	has at least 1 treated friend	0.00	0.00	0.8660
	% matched friends in school	0.03	0.03	0.9598
	% matched friends baseline dropouts	0.00	0.02	0.3931
Change in specific identities of friends between R1 and R2	# friends dropped in R2	3.13	3.23	0.0877
	# friends added in R2	3.28	3.33	0.3416
	# matched friends dropped in R2	0.54	0.49	0.1923
	# matched friends added in R2	0.42	0.52	0.0009
	# treated friends dropped in R2	0.06	0.62	0.0000
	# treated friends added in R2	0.06	0.75	0.0000
	# treated SG dropped in R2	0.00	0.03	0.0000
	# treated SG added in R2	0.01	0.06	0.0000
	# treated dropouts dropped in R2	0.03	0.28	0.0000
	# treated dropouts added in R2	0.02	0.31	0.0000

TABLE 12. Cross-EA Spatial Spillover Effects.

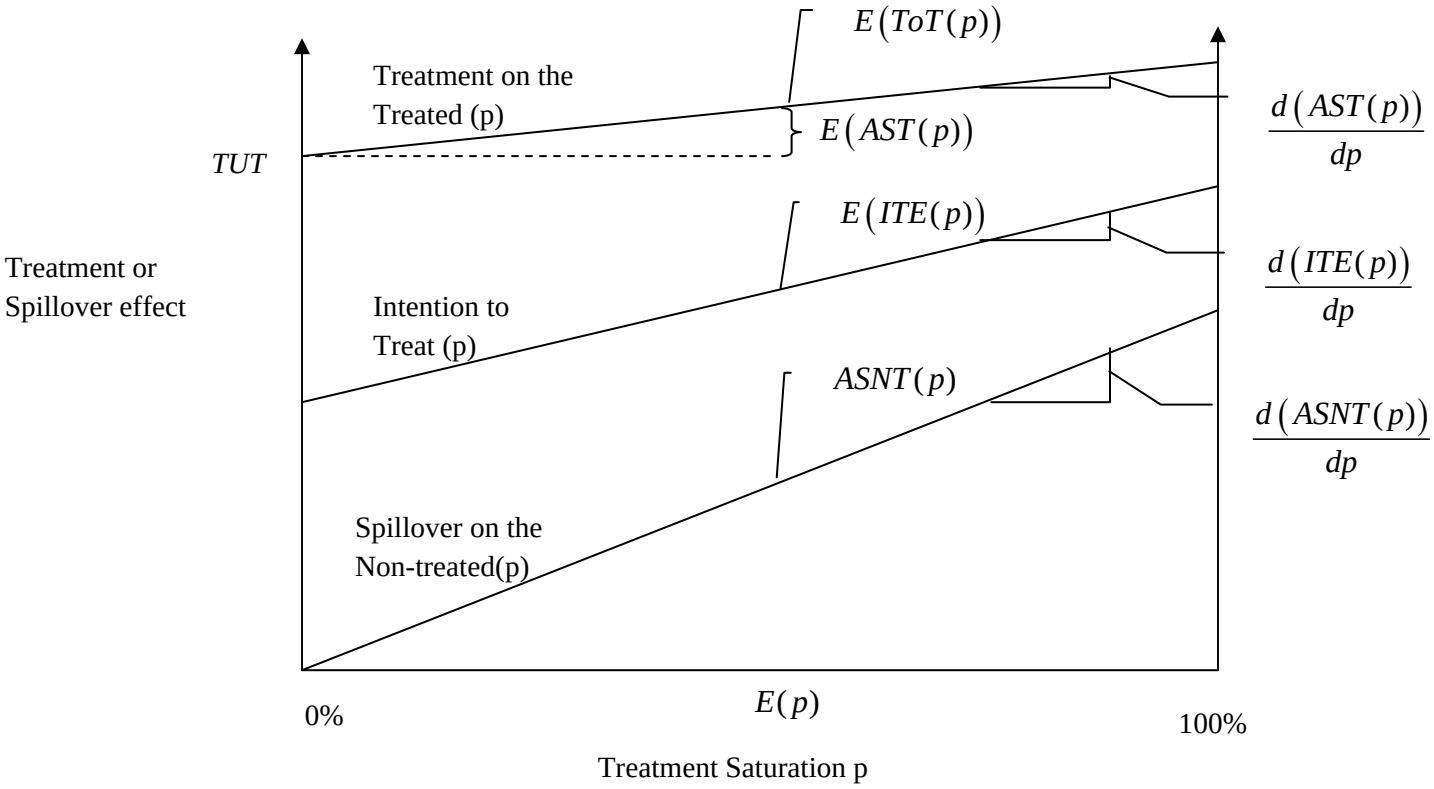
Robustness check using cross-EA variation in treatment intensity

	Dependent Variable:															
	Terms Enrolled		English Test Scores		Cognitive Test Scores		Ever Married		Ever Pregnant		HIV Positive		Expenditures on girls' own consumption		Money spent by girl on other girls' consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
CCT	0.119	0.119	0.067	0.020	0.136	0.048	-0.006	-0.024	0.022	-0.006	0.011	0.007	2.114	0.388	0.147	-0.120
	(0.042)***	(0.083)	(0.052)	(0.085)	(0.049)***	(0.088)	(0.021)	(0.042)	(0.025)	(0.039)	(0.011)	(0.016)	(0.865)**	(1.157)	(0.085)*	(0.088)
UCT	0.036	0.020	-0.032	-0.119	0.089	-0.040	-0.058	-0.077	-0.052	-0.079	-0.017	-0.018	0.518	-1.104	-0.190	-0.465
	(0.050)	(0.110)	(0.070)	(0.096)	(0.079)	(0.109)	(0.027)**	(0.046)*	(0.024)**	(0.048)	(0.008)**	(0.020)	(0.970)	(1.380)	(0.054)***	(0.149)***
Within CCT EA Control	0.012	0.015	-0.005	0.005	0.181	0.191	0.009	0.009	0.001	0.000	0.000	0.001	0.268	0.285	-0.078	-0.069
	(0.047)	(0.047)	(0.082)	(0.082)	(0.068)***	(0.068)***	(0.023)	(0.024)	(0.024)	(0.024)	(0.009)	(0.009)	(0.774)	(0.788)	(0.058)	(0.058)
Within UCT EA Control	-0.100	-0.094	-0.085	-0.064	-0.198	-0.171	0.010	0.010	-0.007	-0.010	0.033	0.034	-1.407	-1.270	-0.154	-0.115
	(0.063)	(0.066)	(0.147)	(0.145)	(0.135)	(0.128)	(0.035)	(0.037)	(0.032)	(0.034)	(0.026)	(0.026)	(0.801)*	(0.833)	(0.067)**	(0.061)*
# of treated EAs within 3 km	-0.024	-0.022	-0.020	-0.008	-0.044	-0.023	0.007	0.005	0.007	0.004	0.003	0.004	-0.133	-0.089	-0.062	-0.028
	(0.018)	(0.020)	(0.031)	(0.038)	(0.025)*	(0.030)	(0.010)	(0.012)	(0.010)	(0.011)	(0.004)	(0.005)	(0.311)	(0.334)	(0.030)**	(0.034)
# of treated EAs between 3 & 6 km	0.005	0.015	0.001	0.021	0.001	0.017	-0.001	0.000	-0.001	-0.002	-0.002	-0.003	-0.838	-0.907	-0.003	-0.011
	(0.013)	(0.016)	(0.019)	(0.023)	(0.015)	(0.022)	(0.006)	(0.007)	(0.006)	(0.008)	(0.003)	(0.003)	(0.252)***	(0.275)***	(0.020)	(0.019)
# of total EAs within 3 km	0.015	0.013	0.025	0.015	0.042	0.030	-0.005	-0.004	-0.002	0.000	-0.001	-0.002	0.123	0.037	0.033	0.010
	(0.012)	(0.013)	(0.018)	(0.022)	(0.015)***	(0.018)	(0.006)	(0.007)	(0.006)	(0.007)	(0.003)	(0.003)	(0.206)	(0.231)	(0.017)**	(0.021)
# of total EAs between 3 & 6 km	-0.002	-0.006	-0.008	-0.017	-0.007	-0.015	-0.001	-0.002	0.002	0.002	0.000	0.001	0.353	0.389	0.000	0.002
	(0.007)	(0.008)	(0.011)	(0.014)	(0.008)	(0.012)	(0.003)	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)	(0.139)**	(0.137)***	(0.011)	(0.011)
Treated individual * # of treated EAs		0.002		-0.006		-0.061		0.019		0.021		0.000		0.746		-0.057
within 3 kilometers		(0.039)		(0.051)		(0.060)		(0.022)		(0.023)		(0.011)		(0.799)		(0.065)
Treated individual * # of treated EAs		-0.029		-0.066		-0.051		-0.003		0.005		0.004		0.197		0.022
between 3 and 6 kilometers		(0.024)		(0.025)***		(0.032)		(0.014)		(0.013)		(0.005)		(0.399)		(0.050)
Treated individual * # of total EAs		0.004		0.020		0.039		-0.008		-0.011		0.002		-0.165		0.042
within 3 kilometers		(0.021)		(0.027)		(0.030)		(0.011)		(0.012)		(0.006)		(0.362)		(0.036)
Treated individual * # of total EAs		0.012		0.028		0.027		0.002		0.000		-0.003		-0.079		-0.001
between3 and 6 kilometers		(0.013)		(0.014)**		(0.018)		(0.007)		(0.007)		(0.003)		(0.221)		(0.027)
Observations	2,579	2,579	2,612	2,612	2,612	2,612	2,649	2,649	2,650	2,650	2,572	2,572	2,649	2,649	2,649	2,649
R-squared	0.1	0.1	0.41	0.41	0.18	0.19	0.14	0.14	0.19	0.19	0.03	0.03	0.11	0.11	0.03	0.03

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation of one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

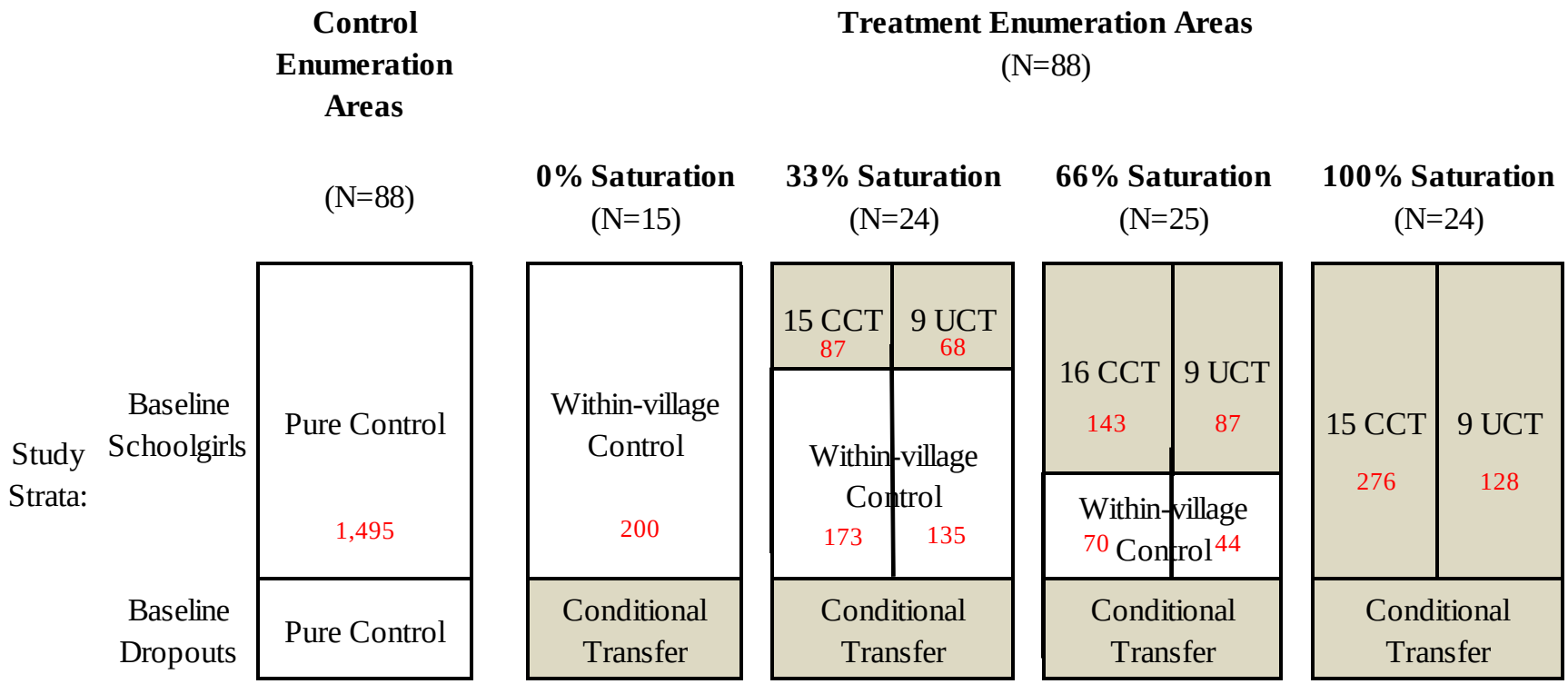
FIGURES.

Figure 1. Example of Spillover Effects.



Graphic demonstrates an example of a positive spillover effect that is stronger for untreated units than treated units. Compliance is assumed to be .5, so that the Intention to Treat lies halfway between the outcome for compliers and non-compliers.

Figure 2. Spillover Research Design.



Shaded cells indicate treatments.

Red numbers give sample sizes at the individual level per cell.

Household transfer amounts randomized at the EA level, monthly values of $4, $6, $8, $10.

Participant transfer amounts randomized at the indvidual level, monthly values of $1, $2, $3, $4, $5.

Figure 3. Map of EA-level Treatment Status.

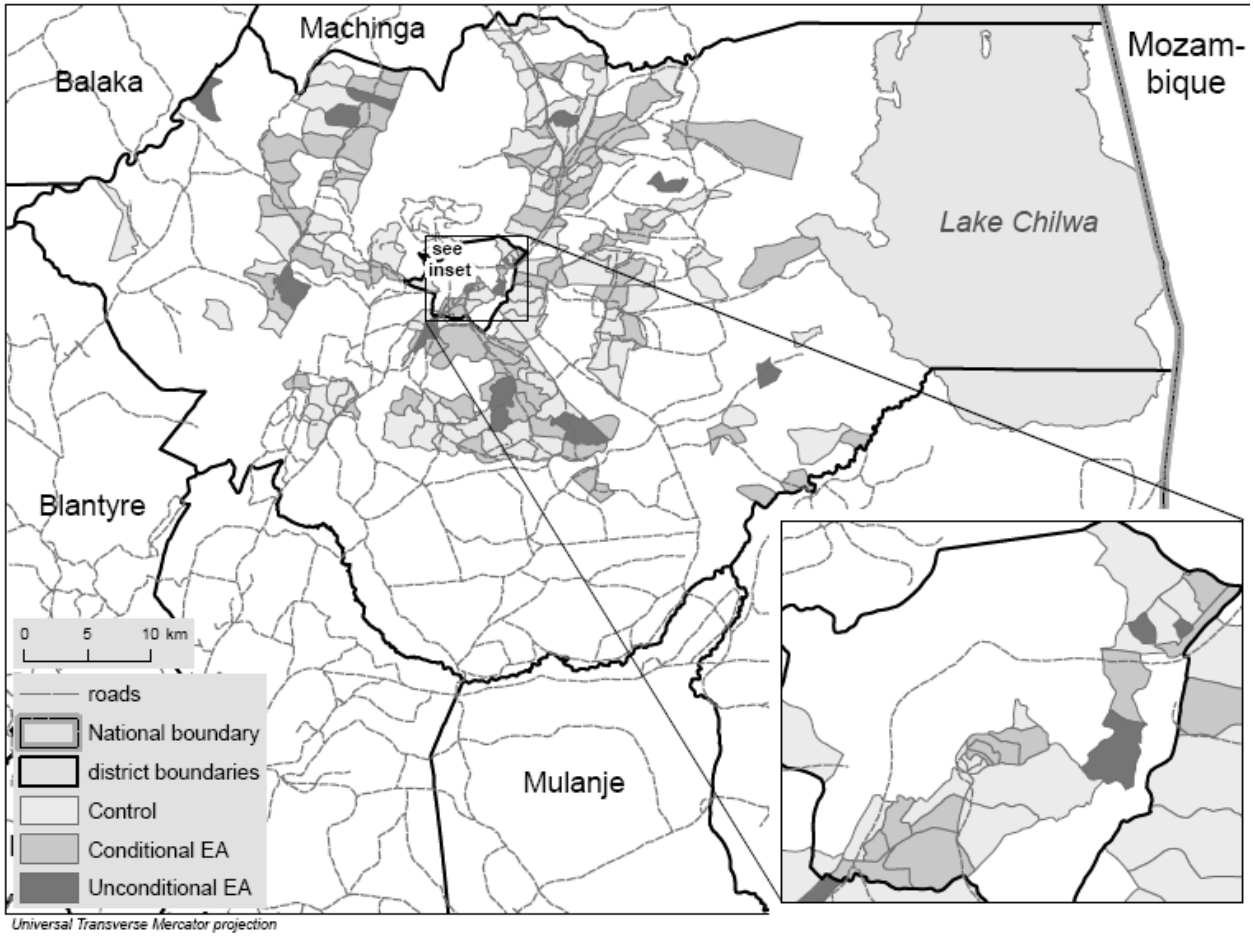


Figure 4. The Empirical Distribution of Saturations.

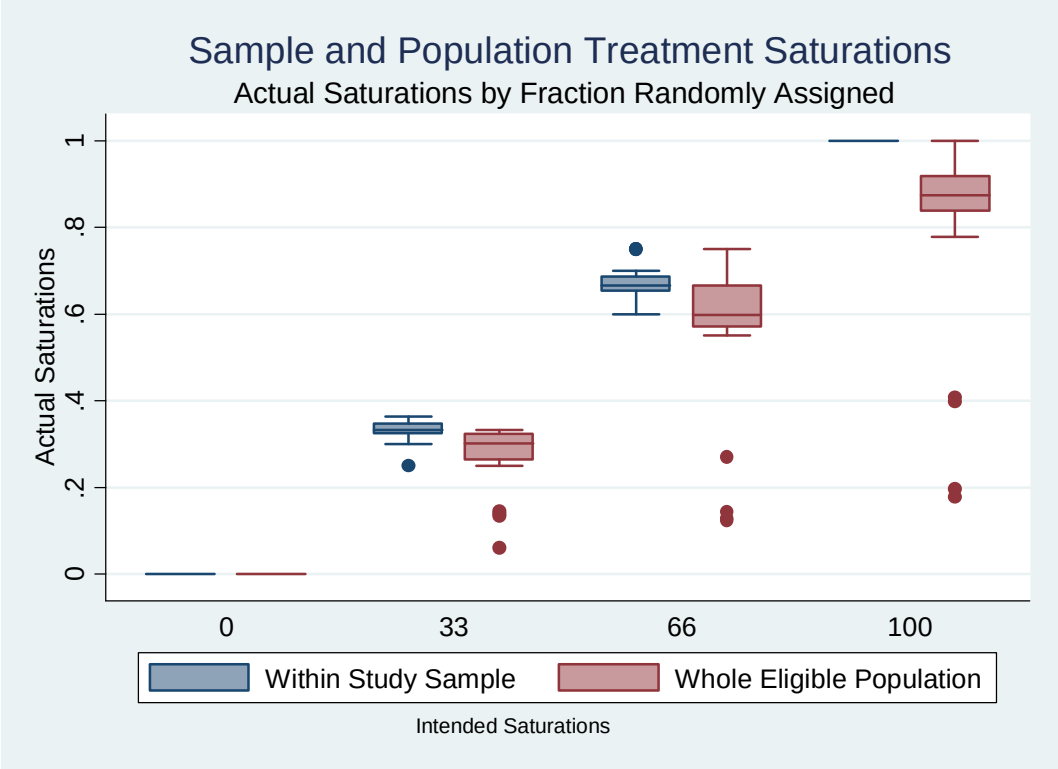


Figure 5. Saturation Impacts on Enrollment.

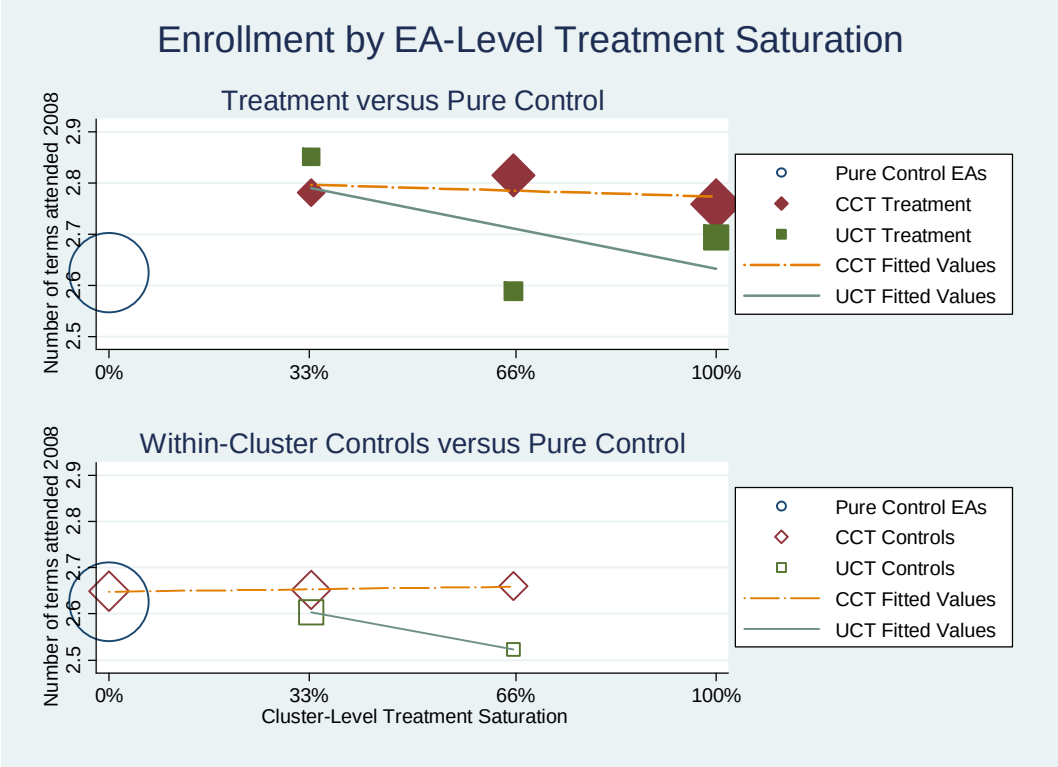


Figure 6. Saturation Impacts on Cognitive Test Scores.

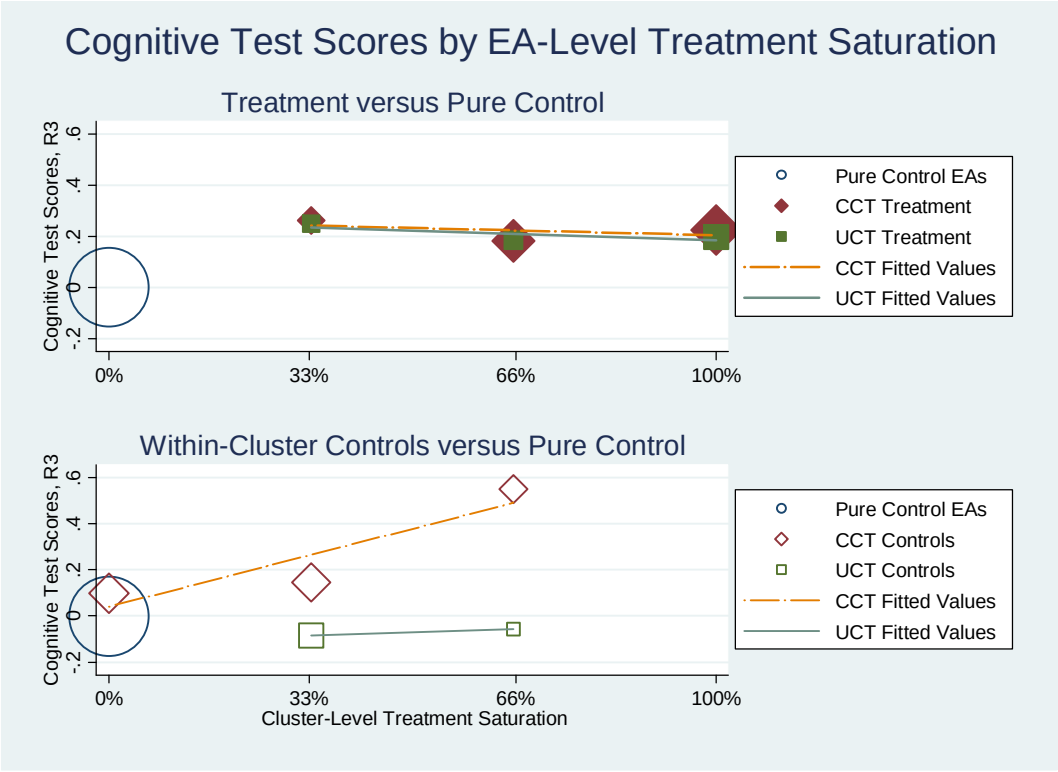


Figure 7. Saturation Impacts on Pregnancy.

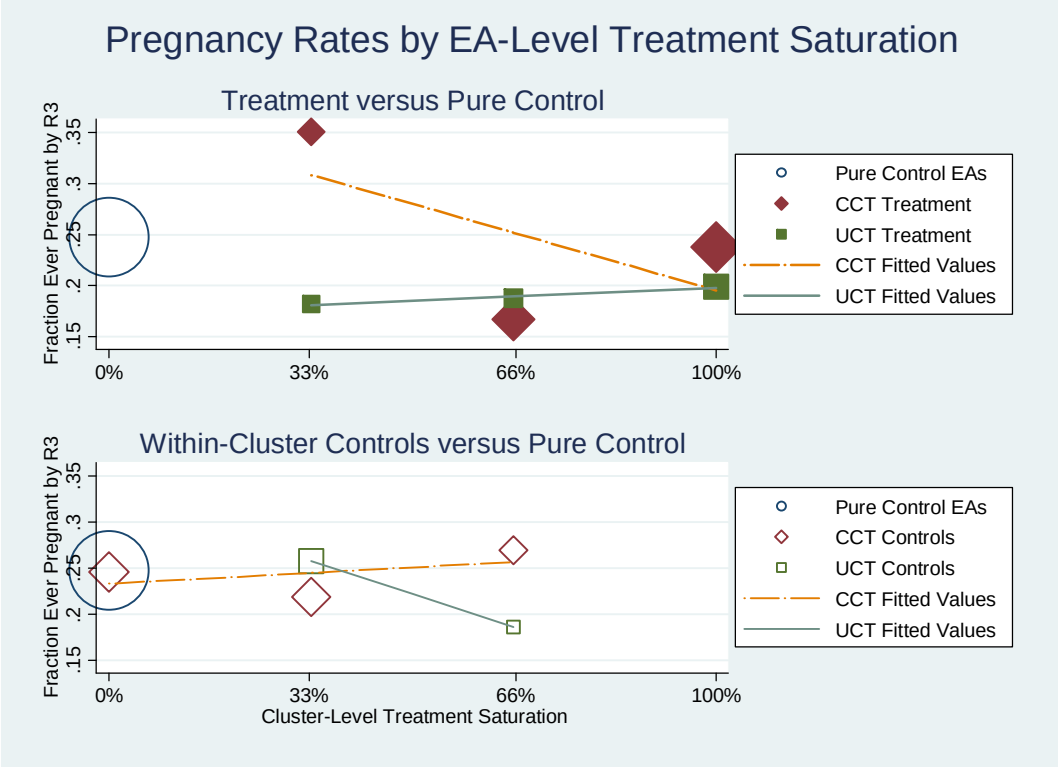


Figure 8. Saturation Impacts on Beneficiary Consumption in CCT EAs.

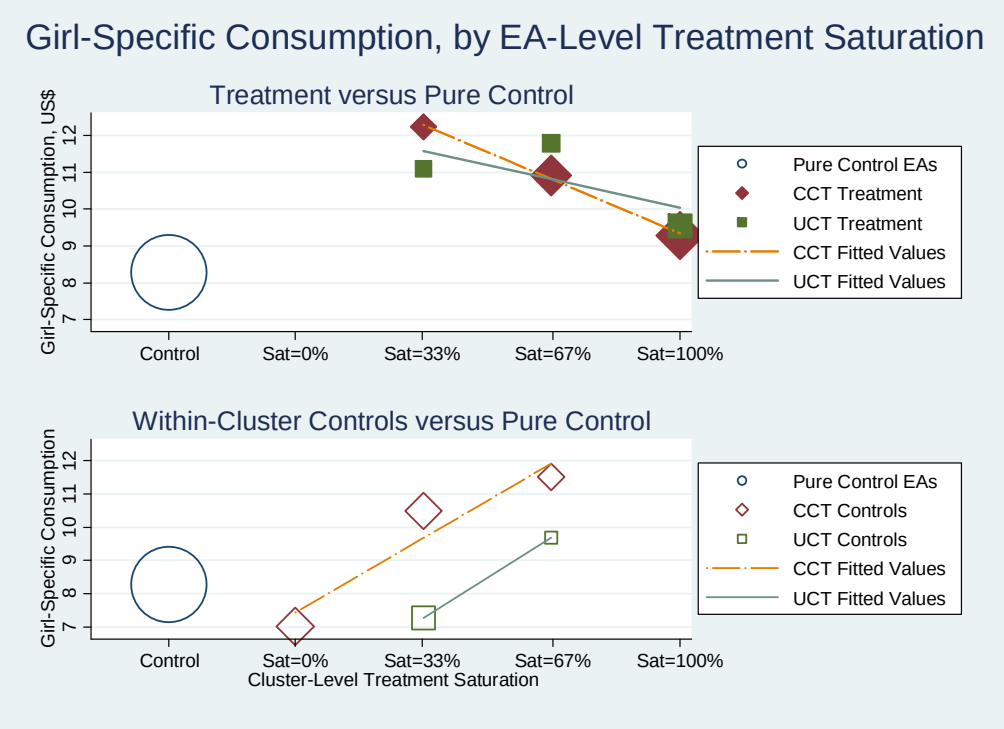


Figure 9. Comparison of Research Designs on Social Network Treatment Saturation.