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

February 24, 2012

Abstract

This paper formalizes the design of experiments intended specifically to study spillover effects. By first randomizing the intensity of treatment within clusters 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 rich set of treatment effects from such an experiment, and provide formulae for power calculations with two-level randomization. The technique is implemented using a Cash Transfer program in Malawi; we find spillover effects to be relatively muted at the cluster level but more potent within households. Conditional cash transfers appear to exert a positive educational spillover in treated villages, and we find no evidence that the protective effect the program exerts on HIV/AIDS is generated by diverting harm onto others.

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

Thanks to useful comments from seminar participants at Caltech, Paris School of Economics, Stanford, and the University of Colorado. Corresponding Author: McIntosh, 9500 Gilman Drive, La Jolla CA, 92093-0519. ctmcintosh@ucsd.edu.

1 Introduction

Spillover effects have traditionally been seen as the Achilles' heel of randomized controlled trials (RCTs). The counterfactual notation used to write down experimental estimands imposes the stable unit treatment value assumption (SUTVA), and the possibility of interference between units represents a major threat to the internal validity of many standard research designs. Recognizing this, a great deal of empirical work has emerged in the past decade trying to relax the assumptions over cross-unit contamination under which RCTs are conducted, and to bring spillovers under the lens of experiments. This has included work to uncover network effects using experimental variation across treatment clusters (Miguel and Kremer 2004, Bobba and Gignoux 2010), using within-cluster units that remained untreated (Duflo and Saez 2003, Barrera-Osoria et al. 2008, Angelucci and De Giorgi 2009, Bobonis and Finan 2009, Lalive and Cattaneo 2009, Alix-Garcia et al. 2011), exploiting plausibly exogenous variation in within-network treatments (Duflo and Saez 2002, Munshi 2003, Conley and Udry 2010, Beaman 2010, Babcock and Hartman 2010), or intersecting an experiment with pre-existing networks (Macours and Vakis 2008, Oster and Thornton 2009, Chen et al. 2010).

In the past few years, researchers have begun to inject experimental variation directly into the intensity of spillover effects by varying the saturation of individuals treated within treated clusters.[1] This paper provides a general setting in which to consider the treatment effects that can be recovered directly from randomized saturation experiments, as well as laying out assumptions under which we can estimate a rich set of saturation-sepecific treatment and spillover effects in programs with imperfect compliance. We then provide power calculation formulae for two-level randomization that nest the standard blocked and clustered designs as special cases. This formalization provides guidance in thinking through optimal research design as a function of some standard quantities (sample size and intracluster correlations in outcomes) and some novel ones (the anticipated magnitude of spillover effects and intracluster correlations in treatment status). These design choices hinge fundamentally on Manski's (1993) reflection problem in ways that cannot be resolved without special experimental design.

We then present an implementation of the technique in which the fraction of eligible girls receiving

[1] Applications of this idea include studies of the spillovers from Get Out The Vote campaigns where the intensity of treatment is randomized spatially (Sinclair et al. 2011, Gine and Mansuri 2011), and the displacement effect caused by job training programs when the fraction of jobseekers treated is randomized by city (Crepon et al. 2011).

a cash transfer in Malawi was randomized at the village level. The program provided transfers to adolescent schoolgirls in Malawi, and was designed to test whether regular cash payments averaging $10 per month could decrease HIV infection. In previous work, we have detailed that the program's conditional arm (CCT) improved schooling outcomes while the unconditional arm (UCT) delayed marriage and fertility (Baird et al., 2011), and that both arms decreased HIV while the program was in operation (Baird et al, 2012). In the context of a program that appears to protect girls from high-risk partners via economic empowerment, an obvious concern is displacement: is any beneficial effect simply causing harm to be diverted towards other girls? This concern was paramount during the design phase of the project, and hence we implemented an experiment that varied treatment intensities directly. We use the rich variation in spillover effects generated by our experiment to show that this displacement effect did not occur, and that if anything the program conferred a protective effect against HIV infection on the sisters of treated girls that was still substantial six months after the program had ended.

We examine three sets of outcomes: academic (enrollment and cognitive scores), demographic (pregnancy and HIV), and economic (household and beneficiary-specific consumption). We present results showing separate spillover effects for the two sub-arms of the transfer experiment: transfer conditional on school attendance and transfers given unconditionally. 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 actual social networks of our study subjects had we used a blocked, clustered, or randomized saturation method to assign treatment. We exploit the experimental variation in treatment saturation within households and social networks to look for evidence of spillover effects within these more tightly defined social groups.

Each of these sets of outcomes has a literature suggesting that substantial spillover effects may exist. 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 tendency of HIV to spread through social and geographic networks in Sub-Saharan Africa is well-documented (Obbo 1993, Behrman et al. 2003). The cash transfers themselves can be an additional source of spillover effects; Angelucci et al. (2010) argue that intrahousehold transfers will modulate the impact of cash transfer programs by allowing extended families to push resources towards secondary-school children who might otherwise have dropped out.

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 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.

Our results indicate that the spillover effects of this Malawian transfer program were quite muted overall. We see an improvement in cognitive test scores among the within-village CCT control just as large as in 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. 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. 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. When we move to the analysis of spillovers within other networks, we find no evidence that outcomes are being transmitted between friends, but siblings appear to confer important spillovers on each other. Both CCT and UCT programs have some positive educational spillovers on siblings of treated respondents, and HIV prevalence falls by 3 and 7 percentage points for each CCT and UCT-treated sister.

Section 2 of the paper 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 highlighting 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 C 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 pure control. While it is typically logistically more straightforward to treat at the cluster than the individual level, and 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 high intra-cluster correlations might otherwise predispose us to pursue a blocked design, it turns out to be critical that we understand *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 central for research design, because endogenous effects will violate SUTVA in the estimation of treatment effects while correlated and contextual effects will not. 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.

This simple but fundamental design choice thereby ends up on the horns of the reflection problem. The research design choice requires us to understand the extent to which intra-cluster correlations are driven by spillover effects, a decomposition which is not generally possible with observational data. While the Intention to Treat effect (ITT) from a clustered design 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. 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.

Perhaps the most-studied example of a partial population experiment is Progresa/Oportunidades in Mexico, which has been used to examine 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 a treatment selection decision at the cluster (village) level as well as an objective poverty eligibility threshold at the household level, and so both eligible and ineligible units in treatment villages can be compared to a pure control group. 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 average spillover effect.

A simple partial-population experiment does not however permit the estimation of two important relationships. First, most 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%). This environment provides no experimental variation for the estimation of the saturation effect. 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. 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 average spillover effect estimated by these limited experiments may be incorrect. Fundamentally, 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.

Secondly, many well-known partial population experiments exploit within-cluster controls that were not treated because they were *ineligible* for the program. This distinction is important because the normal way of using instrumental variables to estimate the ToT is biased if SUTVA is violated, but we provide here an alternate way of backing out ToT effects without requiring SUTVA if we know the spillovers among eligible non-compliers. Spillover effects in the ineligible population may provide a poor measure of the actual spillover observed among those who were eligible. 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.

We now present a theoretical framework intended to capture the variation identified by experiments that vary saturation of treatment. The explicit purpose of conducting such experiments is to estimate consistent treatment and spillover effects in the face of interference between units within a cluster. Manski (2010) works in a completely general framework that permits arbitrary forms of spillovers between units, but we follow Halloran and Struchiner (1995) and Tchetgen and VanderWeele (2010) in placing additional structure on the problem in order to be able to estimate the mean and variance of the treatment and spillover effects generated in randomized saturation experiments.

A first assumptions is required to utilize untreated clusters as a counterfactual in an environment that permits interference between units. This assumption maintains SUTVA across clusters but relaxes it within clusters, and is tested in Section 5.5.

Assumption 1. There is no cross-cluster interference in outcomes.

The next assumption is required in order to be able to use the observed sample to estimate variances in experiments where interference is permitted between units. Under the normal non-interference assumptions, each unit represents an independent draw from the underlying distribution of individual errors (as well as potentially each cluster representing an independent draw from the distribution of clusters). When we permit spillover effects between units in a cluster, we are confronted by the problem that we observe only a singleton of all the potential individual-level configurations of treatment for any given cluster-level assignment. Without further assumptions, we therefore have no information with which to make inference about the variance of these treatment effects. We propose a variant of Hudgens and Halloran's (2008) 'Stratified Interference' assumption, which says that for any treatment status and cluster-level saturation, potential outcomes for a given individual will be independent of the exact identity of the other individuals assigned to treatment and control. This explicitly rules out network-driven heterogeneity in the intensity of spillover effects, an assumption which to which we return in Section 5.

Assumption 2. Consider an individual i located in cluster j, where we first assign a cluster-level saturation of treatment $S_j \in [0,1]$ and then assign individual treatment status $T_{ij} \in 0,1$ conditional on these cluster-level saturations. Individual compliance with treatment is given by $R_{ij} \in 0,1$, and X_{ij} is a vector of covariates. For any fixed cluster-level treatment saturation $S_j = p$, the potential

outcome Y_{ij} is a function of T_{ij} and S_j but is not affected by the specific identities of the individuals assigned to treatment.

These assumptions imply that the potential outcome can be written $Y_{ij}(T_{ij}, R_{ij}, S_j; X_{ij})$, using which we can define the Intention to Treat effect (ITT) conditional on the cluster-level saturation of treatment p:

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

The corresponding term for the average spillover on non-treated units in treated clusters is:

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

A 'spillover effect' is implied if $ASNT(p) \neq 0$ and a 'threshold effect' if $ASNT(p)$ or $ITT(p)$ vary sharply across some part of the distribution of p.

Compliers will experience two types of treatment effects; a direct treatment effect from the program as well as a spillover effect that arises from treatment of agents around them. A natural way of decomposing these two effects is to rewrite the Treatment on the Treated (ToT) by adding and subtracting the outcome for a treated individual who receives no spillover (this would be the 'uniquely treated' individual in a cluster). This decomposes the ToT into a 'Treatment on the Uniquely Treated' and a saturation-dependent spillover effect. Thus,

$$ToT(p) = E(Y \mid T = 1, R = 1, S = p) - E(Y \mid T = 0, R = 1, S = 0).$$

$$= TUT + AST(p),$$

with the 'Treatment on the Uniquely Treated' TUT given by:

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

and an 'Average Spillover on the Treated', given by:

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

Among non-compliers, we have a spillover term which is conceptually similar to the $ASNT(p)$; this is experienced by those who were in treatment clusters but did not themselves receive the treatment.[2] To motivate how spillover effects differ across groups, Figure 1 provides an example of a beneficial, saturation-driven program with a positive spillover effect; it illustrates how the $ITE(p)$, the $ASNT(p)$, and the $ToT(p)$ might vary across the distribution of treatment saturations.

We can now consider the effects recovered by different research designs. A blocked design features a fixed treatment saturation in each cluster and uses within-cluster controls as counterfactuals. A clustered design features treatment saturations that equal either $\bar{p}$ or zero, but no data is gathered on untreated individuals in treatment clusters. Clustered impacts are estimated by comparing treatment and control clusters.

In the face of the SUTVA violations modeled here, the blocked design proves highly unattractive. 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 for sample compliance rate r the ITT recovered by the blocked design will be:[3]

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

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

The blocked estimator is biased by the term $E(Y \mid T = 0, S = \bar{p}) - E(Y \mid T = 0, S = 0) = ASNT(\bar{p})$.

The clustered design, on the other hand, recovers the correct ITT as long as SUTVA holds between clusters:

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

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

The blocked design is biased by the expected value of the spillovers in the counterfactual, and the blocking leaves no variation in the distribution of p that can be used to investigate this bias term.

[2] The Average Spillover effect on Non-Compliers is $ASNC(p) = E(Y \mid T = 1, R = 0, S = p) - E(Y \mid T = 0, = 0, S = 0)$.

[3] To simplify these expressions we maintain Assumption 3, that $ASNC(p) = ASNT(p)$.

The clustered design is unable to estimate how the treatment or spillover effects vary with saturations (because these are fixed in a typical clustered design), and consequently it also cannot differentiate the $AST(p)$ from $ASNT(p)$, or the $AST(p)$ 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 $ToT(\bar{p}) \neq \frac{ITE(\bar{p})}{r}$ unless $ASNT(\bar{p}) = ASNC(\bar{p}) = 0$. Blocked designs, therefore, are at the mercy of spillovers, and clustered designs provide the correct answer but no way to investigate spillovers directly.

A partial population design is a third alternative, providing direct evidence as to the nature of spillover effects. Similar to a clustered design, partial population experiments feature pure control clusters where $S = 0$ as well as partially treated clusters where $S = p < 1$. Critically, in a partial population data is also gathered on untreated units in treated clusters. Such designs provide an experimental estimate of the $ASNT(p)$ by comparing the within-village controls to the pure controls. While partial population experiments provide direct evidence on spillover effects, they cannot shed light on saturation or threshold effects because the treatment saturations in such experiments are either fixed or endogenous.

The design proposed here is the randomized saturation design: some clusters are left untreated as controls, then the saturation of treatment is randomly assigned in treatment clusters, and then individual treatments are randomly assigned according to the cluster-level saturation. This design has several unique advantages in the analysis of saturation effects, the most obvious of which is direct experimental estimate of the saturation-contingent $ITT(p)$ and $ASNT(p)$. These can be estimated non-parametrically by taking the difference between treated (ITT) or within-village control ($ASNT$) observations and the pure control across the distribution of p.

We can give the regression estimators of these quantities by taking the entire sample, 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 $E(ASNT(p)) = \eta_2$.

A second potential advantage of the randomized saturation design emerges when we consider estimation of the ToT in a study with non-compliance in which interference is permitted. With interference, the exclusion restriction is violated when we try to instrument for R_{ij} with T_{ij}, and so the standard method of recovering the ToT is not available. Unlike many partial-population experiments in which the within-village controls are *ineligible* for the treatment, in a randomized saturation design the within-village controls come from the same population as the treatment sample. Each treatment village thus contains two types of untreated individuals, the within-village controls and the non-compliers, with the former group being an unselected and the latter a selected sample from the same population. If we are willing to assume that the spillover effects within these two untreated samples are similar, then we can algebraically back out an estimate of the $ToT(p)$ even in the face of interference. [4]

Assumption 3. $ASNT(p) = ASNC(p)$.

This assumption is not required to estimate the $ITT(p)$ or $ATNT(p)$, but provides a way of replacing the IV estimator's implicit assumption that $ASNC(p) = ASNT(p) = 0$ with an estimate the magnitude of this spillover effect. In the most widely examined partial population experiments (such as Progresa/Oportunidades) the within-village controls are untreated because they are ineligible for the program, making it difficult to argue that the $ASNT(p)$ provides a counterfactual for the $ASNC(p)$. If we maintain Assumption 3, we can recover an estimate of the Treatment on the Treated for each value of p:

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

A parsimonious linearization 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 $\frac{d(ASNT(p))}{dp}$. If in addition

[4] In principle, one could test this assumption by invoking selection on observables and seeing whether outcomes among the sample of non-compliers are similar to those among a matched sample of within-village controls. Since this would require subjecting experimental data to a completely observational counterfactual, we prefer to maintain the similarity of spillover effects as an assumption.

we maintain Assumption 3, we can estimate $\frac{d(AST(p))}{dp} = \frac{\beta_3-(1-r)\beta_4}{r}$, 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 $p = 0$, $TUT = \frac{\beta_1}{r}$ 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. From (1) we can back out $E(ToT(p)) = \frac{(\eta_1-(1-r)\eta_2)}{r}$, and we can combine estimates from (1) and (2) to get $E(AST(p)) = E(ToT(p)) - TUT = \frac{\eta_1-(1-r)\eta_2-\beta_1}{r}$. 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 using F-tests.

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:

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

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 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

We now derive expressions that can be used to calculate the power of two-level randomization 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:

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

The design problem is how to assign a treatment T_{ic} across a given C clusters with n observations in each cluster, for a total of $N = nC$ 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 term is the intra-cluster correlation of the treatment itself, and gives the key expectation that allows for power calculation with a two-stage randomization. This cluster-level 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.

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* $\pi \sim (\mu, \eta^2)$*. 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*

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

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. For any variance, $\mu = 0.5$ minimizes $SE\left(\widehat{\beta}_1\right)$.

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

The formula in Theorem 1 nests the familiar expressions for the standard error of the blocked and clustered designs:

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:[5]

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

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:

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

The expression in Theorem 1 also provides intuition for two fundamental results. First, for the well-known fact that the blocked design maximizes power in the face of intra-cluster correlation. For any given μ, the MDE is minimized when $\rho = 0$, meaning that there is no variation in treatment saturations across clusters. This effect occurs only because of the presence of intra-cluster

[5]Note that 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 $C \to \infty$.

correlations in the underlying dependent variable. Following from this, if there is no intra-cluster correlation in the dependent variable, then the MDE is independent of ρ and the research design is irrelevant to the power of the study.

Randomized Saturation design:

Having framed the blocked and clustered designs as the extreme points on the possible values of ρ, we can consider the power implications of the 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.

Theorem 2 *Suppose that we compare power of the three designs holding μ constant and using within-village controls as counterfactuals. 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 $MDE^B < MDE^{RS} < MDE^C$.* ■

An extension of this argument can be used to show the the randomized saturation design will have lower power than the partial population design, holding constant the fraction of clusters treated and the overall treatment saturation. If we conducted a partial population experiment by treating half of the individuals in half of the clusters, then within treated clusters we have a blocked design. Varying the treatment saturation around 50% within these clusters increases the intra-cluster correlation in treatment status ρ. Thus any move in the direction of randomizing saturations results in a power loss.

2.3.2 Using only pure controls as counterfactuals:

The treatment of power up to this point has made the canonical assumption that the unexplained clustering in the data arises from a common unobservable shock. In that setting, within-cluster counterfactuals are particularly useful because they allow for estimation of treatment effects netting out these shocks. Manski's endogenous effects, on the other hand, can lead to outcomes that appear clustered because we have failed to account for the lack of dependence between units. In this case,

within-village controls are potentially polluted by spillover effects, and thus only comparison to pure control clusters provides an unbiased counterfactual.

If we eschew the use of within-village controls, then varying the treatment saturation within treated clusters varies the number of usable observations per treatment cluster. 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.

Theorem 3 *Suppose we use two-stage randomization procedure as described in Theorem 1, but do not include the within-village controls in the counterfactual estimate. Then the standard error of the estimated treatment effect is*

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

Where $\phi = \frac{\sum_{c=1}^{C} n_c^2}{N}$ *is the average of* n_c^2.

Proof. See Appendix. ■

We see that ϕ is minimized at $n_c = n$ for all c, at which point $\phi = n$. Variation in the treatment saturations now passes directly into the number of observations per treated cluster, and hence causes an additional power loss. The intuition for this is most clearly seen by considering that the power curve for the number of observations per cluster is concave, and so by Jensen's inequality an experiment that varies saturation around some mean will have a lower overall average power than if all clusters were treated with the mean saturation. Interestingly, it may now be optimal to have $\mu > 1/2$, as the dependence of ϕ on μ can offset some of the power loss from varying n_c among the treated clusters.

The expression for the power to detect a spillover effect is the same as that in Theorem 3, with the size of the within-village control sample replacing the size of the treatment sample. Indeed, if the average treatment saturation of treatment in treated clusters is .5, then the power to detect treatment and spillover effects will be identical. If we think that spillovers will be large, we can detect them with a smaller within-village control and leave more power for the study of treatment effects. If we expect intra-cluster correlations in outcomes to be large, we need to retain a larger number of pure control clusters, but particularly with a low ICC it may be optimal to have fewer than half of the clusters serve as pure controls. Finally, if we expect the saturation impacts to be

large then we can detect them with less saturation variation, thereby improving the power of the overall study via ϕ.

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 lies between. If the true relationship with the saturation is linear then the two designs should be equivalent.

In conclusion, the cost at which we buy the ability to test more nuanced forms of spillovers by randomizing saturation is power. 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.

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, just after the program ended) to analyze impacts.

We conducted three additional forms of data collection to provide independent outcome mea-

sures 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. 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 2,658 sampled baseline schoolgirls on whom we have three rounds of panel data, 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 experiment within the sample of baseline schoolgirls. The fact that the saturations were done within the sample rather than the population causes the actual saturation of treatment to be lower than apparent 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 a specification that modifies Equations (1) and (2) to allow conditional and unconditional treatments to have separate treatment and spillover effects. 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

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

and

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

We examine all baseline schoolgirls in the last survey wave, controlling for a basic set of covariates

and clustering standard errors at the EA level. Table 3 presents two sets of results for each outcome, first showing simple ITT and ASNT effects by estimating Equation (4) in the odd-numbered columns, and then proceeding to test for the presence of saturation effects using Equation (5) in the even-numbered columns. The treatment effects of the program have been documented 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. The cluster-level spillover effect appear quite muted despite the relatively stark direct treatment effects. The sole $E(ASNT)$ that proves to be significant using (4) is the cognitive functioning of the CCT within-village controls; here we see a substantial 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, we find that the cognitive spillovers of the CCT to the within-village controls are sensitive to saturations, with higher treatment fractions leading to larger spillovers. Within-village controls have negative slopes on the saturation coefficients for both educational outcomes, suggesting the possibility of a congestion effect on these outcomes. 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. Expenditures also appear to have saturation-driven effects, with the spillover sample consuming more in UCT EAs where a higher fraction of the sample was treated, both at the household and the beneficiary-specific level.[6] The saturation effects on expenditures for the CCT treatments also suggest that the transfers had a greater tendency to be concentrated on beneficiaries when only a few girls were treated. There is no evidence that the protective effect of the treatment in terms of pregnancy and HIV is generating a negative displacement effect on the within-village

[6]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.

controls.

The bottom two panels of Table 3 explicitly calculate each of the relevant treatment effects for the conditional and unconditional arm, invoking the assumptions laid out in Section 2.3, and using the treatment-specific compliance rate r of .99 for the UCT and .605 for the CCT. 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 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 divide the first row in the even-numbered columns by the compliance rate to calculate the TUT. We can then use the compliance rates to calculate $E(ToT)$, and given these last two quantities we then have $E(AST) = E(ToT) - TUT$. We calculate F-tests on each of these linear combinations of regression parameters to provide significance for these estimates.

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, in the conditional and unconditional arms separately. 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 4 estimates the split counterpart to Equation (3):

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

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.

In Table 4 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 5-6 that provide the visual analog to these regressions, with the only difference being that Table 4 includes covariates as controls while the figures present the raw averages of outcomes within each treatment cell. Again, the table and 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, and household consumption increases only when the treatment saturation is not 100%.

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 networks more concentrated than the cluster: households and peer-group social networks. 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, confirming that the cross-cluster SUTVA assumptions invoked throughout this paper are largely justified by the data.

5.1 Analysis of Spillovers within Households

We can exploit the fact that the sample contains 517 individuals who come from households with more than one sampled girl 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.

Table 5 conducts the household-level spillover analysis. When we look in this concentrated network where motivational effects are likely to dominate, unconditional 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. 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 5 shows no improvement in cognitive performance among the sisters of treated girls, however.

Households appear to generate stronger beneficial spillovers of the program on HIV. Table 5 shows that when two sisters receive CCT treatment together, pregnancy decreases 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 5 are all as we might expect (treatment increases the amount spent on consumption of treated and untreated girls, they are insignificant. Hence, while we might expect the expenditure spillover effects to be concentrated between sisters within households, this appears not to be the case.

As a part of our survey, we collected information on the schooling and marital status of all of the siblings of girls who were a part of the study. These data provide a fascinating additional set of perspectives on spillover effects. By eliminating all studied girls from the sample of siblings and then calculating enrollment and marriage rates, we can study the spillover effect of the program on unstudied siblings. This sample includes all brothers, and thereby provides our only view of how this female-centric program has affected boys. Sisters may be in the unstudied sample either because

they lay outside of the 13-22 age frame used for sampling, or because they are part of the 37.5% of eligibles who were not sampled. By examining whether siblings with treated core respondents differ from those with pure control respondents, we measure spillover effects on to unstudied siblings. Comparing the siblings of within-village controls to the siblings of pure controls gives a measure of a very indirect form of spillover effect, present in the entirety of treatment villages even in households where no-one was treated.

Table 6 carries out this analysis, restricting the analysis to siblings between the ages of 8 and 25. While these household-average outcomes are observable only for core respondents who have siblings, they capture averages among the entire set of siblings rather than just among those who were sampled to the survey, as in Table 5. Thus every treated girl creates a non-zero household treatment saturation in this analysis, while in Table 5 only treated girls with sisters in the sample create non-zero saturations.

These results confirm a weak positive halo effect of CCT programs on educational outcomes. Not only do siblings of CCT-treated respondents have a 4.7 percentage point higher schooling rate (off a base of 66%), but siblings of the within-village controls in CCT EAs are elevated by an identical amount. This implies that the ambient spillover for those in the same village as conditional treatment was the same as the spillover passing through the household, and is consistent both with the EA-level spillovers in Table 3 and with the *lack* of evidence that the household is a powerful conduit for these spillovers in Table 5. Girls and boys experience this cluster-level CCT spillover in similar ways.

UCTs, on the other hand, appear to much less advantage when we consider unstudied siblings. By comparing Tables 5 and 6, it appears that the benefits of UCT programs for education were concentrated on *studied* girls, in part by diverting benefit away from *unstudied* girls. This is the sole evidence of negative spillover effects uncovered in this analysis, and raises the possibility that the overall educational benefits of a UCT program may be smaller than those detected in the sample of studied girls.

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 7 estimates spillover effects using variation in treatment status of the friends. The beneficial network spillovers of the treatment on enrollment appear to 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 protective for friends under both treatments, and while the 5 percentage point decrease in prevalence for each UCT treated friend is large, it is not quite significant. Household expenditures are sharply lower for both CCT treated girls and for their friends as treatment saturations rise, indicating as in the previous section that the pressure to redistribute transfer money increases as the treatment saturation increases. The fact that friends' household expenditures also drop indicates that this money is being distributed through networks other than the social networks of the girls themselves.

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 Bangladesh 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 8 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 8 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 The Benefits of Randomized Saturations when we use Different Networks

Having followed social networks to study spillover effects, we now use them to illustrate a side benefit of the randomized saturation design: this design increases the saturation variation present in alternate definitions of network, so long as other networks display any spatial correlation.

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 7 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 7 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.

5.5 Robustness: Spatial Spillovers

We conclude with an analysis of the spatial dimension of spillovers by examining the extent to which outcomes in a cluster are influenced by the treatment status of neighboring clusters. Having maintained SUTVA across clusters, this provides a check on the extent to which our 'pure' control is indeed unadulterated by spillover effects. To examine this possibility, we follow Miguel and Kremer (2004) and Bobba and Gignoux (2010) in looking at intensity of treatment within distance bands around each cluster. We take GIS data on the locations of the EA centroids, and then calculate cluster-level control 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 9 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 of negative cross-cluster spillovers in consumption for girls, but this does not have a simple spatial pattern. In contrast to Bobba and

Gignoux who find large spillover effects of *Progresa*in Mexico but only on treated individuals, we find no evidence that program beneficiaries experience spillovers from adjacent clusters that are any different from untreated individuals (in other words, the AST and the $ASNT$ measured cross-cluster are both zero). We therefore conclude that the cross-cluster imposition of SUTVA is largely warranted by the data.

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 RCT, 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 can also serve to improve the external validity of randomized trials. Small policy trials conducted in partial equilibrium can miss important scale or congestion effects that will accompany the full-scale implementation of a program. 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. 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? 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 provides experimental framework with superior internal and external validity.

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. 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. If anything, enrollment and cognitive functioning appear to improve in treated clusters. Then, the significant decreases in HIV and HSV-2 that were reported in Baird et al. (2012) could be entirely illusory if it were the case that the program was merely diverting dangerous sexual activity. For both pregnancy and HIV status, the coefficient on treatment saturations for the within-village controls is in fact negative, indicating that the protective effect of this intervention did not come about by displacing harm. 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, (2011).

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, (2012).

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.

Chen, Jiehua, Macartan Humphries, and Vijay Modi, "Technology Diffusion and Social Networks: Evidence from a Field Experiment in Uganda." Working paper, 2010.

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).

Halloran, M. Elizabeth and Claudio Struchiner, "Causal Inference in Infectious Diseases," Epidemiology, 6(2), (1995), pp. 142-151.

Hudgens, Michael, and Elizabeth Halloran, "Towards Causal Inference with Interference," Journal of the American Statistical Association, 103(482), (2008), pp. 832-842.

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.

Obbo, Christine. "HIV Transmission Through Social and Geographical Networks in Uganda," Social Science and Medicine, 36(7), pp. 949-955.

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: Evidence from a Randomized Controlled Trial in Orissa," Working Paper (2011).

Tchetgen, Eric J., and Tyler VanderWeele, "On Causal Inference in the Presence of Interference,"

Statistical Methods in Medical Research, 21(1) (2010), pp. 55-75.

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	Household Consumption Aggregate
	(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	2.759
	(0.216)	(0.478)	(0.059)	(0.150)**	(0.027)	(0.038)	(0.031)	(0.008)	(1.226)	(1.864)
UCT	0.275	0.672	-0.001	0.130	-0.005	0.048	-0.016	0.006	0.144	-0.790
	(0.186)	(0.390)*	(0.067)	(0.122)	(0.026)	(0.034)	(0.036)	(0.009)	(0.656)	(1.266)
Within CCT EA Control	-0.211	-0.167	-0.024	0.090	-0.035	0.036	-0.003	0.006	1.820	0.612
	(0.205)	(0.508)	(0.073)	(0.154)	(0.033)	(0.036)	(0.041)	(0.018)	(1.166)	(2.881)
Within UCT EA Control	0.416	0.086	-0.048	0.295	0.067	-0.111	0.103	-0.056	-0.157	-6.438
	(0.528)	(1.337)	(0.206)	(0.596)	(0.110)	(0.163)	(0.071)	(0.054)	(1.540)	(4.794)
EA CCT Treatment Saturation	0.207	1.060	0.106	0.093	0.146	0.036	0.048	-0.005	-1.497	4.154
	(0.452)	(1.223)	(0.167)	(0.388)	(0.073)**	(0.089)	(0.104)	(0.040)	(2.592)	(6.863)
EA UCT Treatment Saturation	-1.182	0.943	0.249	-0.664	-0.118	0.249	-0.176	0.135	-1.697	14.083
	(0.990)	(2.669)	(0.445)	(1.142)	(0.200)	(0.261)	(0.172)	(0.156)	(3.154)	(11.223)
Observations	2,651	2,651	2,651	2,653	2,653	2,648	2,653	2,652	2,653	2,650
R-squared	0	0.01	0	0.01	0	0	0	0	0	0.01

Regressions are OLS using Round 1 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

TABLE 3.

Basic Spillover Analysis, Split

	Dependent Variable:											
	Terms Enrolled		Cognitive Test Scores		Ever Pregnant		HIV Positive		Expenditures on girls' own consumption		Household Consumption Aggregate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CCT	0.134	0.276	0.166	0.264	0.016	0.060	0.009	0.004	1.706	5.828	3.316	10.709
	(0.045)***	(0.121)**	(0.052)***	(0.131)**	(0.027)	(0.079)	(0.012)	(0.030)	(0.988)*	(2.973)*	(1.501)**	(3.563)***
UCT	0.050	0.214	0.117	0.189	-0.054	-0.130	-0.019	-0.013	0.792	2.871	-0.944	0.358
	(0.050)	(0.100)**	(0.081)	(0.292)	(0.023)**	(0.056)**	(0.008)**	(0.014)	(1.027)	(3.783)	(1.329)	(3.776)
Within CCT EA Control	0.017	0.003	0.192	0.037	0.000	-0.007	0.000	0.008	0.562	-1.144	1.326	1.212
	(0.047)	(0.077)	(0.072)***	(0.079)	(0.024)	(0.031)	(0.009)	(0.014)	(0.900)	(0.770)	(1.436)	(1.968)
Within UCT EA Control	-0.080	0.019	-0.151	-0.330	-0.009	0.083	0.032	0.067	-1.011	-4.667	-0.733	-6.613
	(0.066)	(0.205)	(0.138)	(0.397)	(0.030)	(0.079)	(0.026)	(0.068)	(0.925)	(2.213)**	(1.341)	(3.117)**
EA saturation,		-0.181		-0.124		-0.056		0.005		-5.229		-9.407
CCT Treatment		(0.154)		(0.187)		(0.104)		(0.041)		(3.517)		(3.550)***
EA saturation,		-0.217		-0.097		0.102		-0.008		-2.795		-1.763
UCT treatment		(0.134)		(0.346)		(0.065)		(0.014)		(4.111)		(4.410)
EA saturation,		0.054		0.610		0.029		-0.029		6.620		0.449
Within-CCT Control		(0.169)		(0.272)**		(0.085)		(0.032)		(1.629)***		(5.485)
EA saturation,		-0.242		0.439		-0.224		-0.086		8.950		14.366
Within-UCT Control		(0.529)		(0.875)		(0.153)		(0.121)		(4.632)*		(7.591)*
Observations	2,579	2,579	2,612	2,612	2,650	2,650	2,572	2,572	2,649	2,649	2,646	2,646
R-squared	0.1	0.1	0.18	0.19	0.19	0.19	0.03	0.03	0.1	0.11	0.12	0.12
Estimates for CCT::												
E(ITT)	0.134 ***		0.166 ***		0.016		0.009		1.706 *		3.316 **	
E(ToT)	0.210 ***		0.149		0.026		0.015		2.453 *		4.615 **	
TUT	0.456 **		0.436 **		0.099		0.007		9.633 **		17.701 ***	
E(ASNT)	0.017		0.192 ***		0.000		0.000		0.562		1.326	
E(AST)	-0.246		-0.287		-0.073		0.008		-7.180		-13.086 **	
Estimates for UCT::												
E(ITT)	0.050		0.117		-0.054 **		-0.019 **		0.792		-0.944	
E(ToT)	0.051		0.120		-0.054 **		-0.020 **		0.810		-0.946	
TUT	0.216 **		0.191		-0.131 **		-0.013		2.900		0.362	
E(ASNT)	-0.080		-0.151		-0.009		0.032		-1.011		-0.733	
E(AST)	-0.165 *		-0.071		0.077		-0.006		-2.090		-1.308 ***	

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, and an indicator for ever had sex. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

TABLE 4.

Granular Spillover Analysis

	Dependent Variable:					
	Terms Enrolled	Cognitive Test Scores	Ever Pregnant	HIV Positive	Expenditures on girls' own consumption	Household Consumption Aggregate
	(1)	(2)	(3)	(4)	(5)	(6)
CCT 33%	0.180	0.255	0.089	-0.006	4.089	4.218
	(0.085)**	(0.089)***	(0.049)*	(0.014)	(2.331)*	(1.711)**
CCT 66%	0.185	0.157	-0.018	0.019	2.327	7.408
	(0.054)***	(0.051)***	(0.032)	(0.019)	(1.544)	(2.456)***
CCT 100%	0.084	0.151	0.018	0.006	0.607	0.386
	(0.063)	(0.086)*	(0.042)	(0.019)	(1.226)	(1.118)
UCT 33%	0.192	0.118	-0.080	-0.025	0.942	-1.584
	(0.059)***	(0.220)	(0.038)**	(0.008)***	(2.901)	(2.657)
UCT 66%	-0.003	0.191	-0.089	-0.002	2.686	1.438
	(0.084)	(0.088)**	(0.039)**	(0.016)	(1.630)	(2.004)
UCT 100%	0.020	0.072	-0.020	-0.026	-0.435	-2.080
	(0.069)	(0.109)	(0.026)	(0.007)***	(0.818)	(1.471)
Spillover CCT 0%	0.002	0.100	0.008	0.014	-1.420	2.258
	(0.084)	(0.077)	(0.032)	(0.015)	(0.696)**	(2.108)
Spillover CCT 33%	0.026	0.097	-0.036	-0.017	1.750	-0.920
	(0.055)	(0.079)	(0.033)	(0.009)*	(1.425)	(1.258)
Spillover CCT 66%	0.035	0.577	0.046	0.003	2.640	3.679
	(0.072)	(0.163)***	(0.053)	(0.018)	(1.271)**	(3.140)
Spillover UCT 33%	-0.061	-0.185	0.009	0.038	-1.692	-1.863
	(0.063)	(0.158)	(0.036)	(0.032)	(0.990)*	(1.217)
Spillover UCT 66%	-0.141	-0.036	-0.067	0.011	1.299	3.054
	(0.166)	(0.249)	(0.040)*	(0.027)	(1.355)	(2.498)
Observations	2,579	2,612	2,650	2,572	2,649	2,646

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, and an indicator for ever had sex. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

TABLE 5.

Analysis of Household Networks, Treatments Split.

	Terms Enrolled	Cognitive Test Scores	Ever Pregnant	HIV Positive	Money spent by beneficiary on own consumption
VARIABLES	(1)	(2)	(3)	(4)	(5)
CCT Treatment	0.12***	0.08	0.03	-0.01	0.75
	(0.05)	(0.07)	(0.03)	(0.01)	(1.04)
UCT Treatment	0.02	0.18	-0.06**	-0.02**	1.31
	(0.06)	(0.12)	(0.03)	(0.01)	(0.82)
Within-Village Control CCT	0.02	0.18**	0.00	0.00	0.59
	(0.06)	(0.08)	(0.03)	(0.01)	(0.93)
Within-Village Control UCT	-0.07	-0.06	-0.01	0.05	-0.49
	(0.07)	(0.12)	(0.03)	(0.03)	(0.74)
Number of CCT-Treated HH Members	0.09	0.12	-0.05**	0.03	2.35
for Treated Girls	(0.07)	(0.11)	(0.03)	(0.03)	(2.09)
Number of UCT-Treated HH Members	0.22***	0.08	0.02	-0.00	2.19
for Treated Girls	(0.08)	(0.15)	(0.05)	(0.01)	(2.61)
Number of CCT-Treated HH Members	0.05	0.00	-0.02	-0.03***	2.18
for Untreated Girls	(0.15)	(0.24)	(0.07)	(0.01)	(1.81)
Number of UCT-Treated HH Members	0.18*	-0.17	-0.05	-0.07***	-0.06
for Untreated Girls	(0.10)	(0.18)	(0.03)	(0.03)	(1.87)
Number of HH members in Sampling Frame	0.00	0.05	-0.01	0.01*	0.17
	(0.03)	(0.05)	(0.01)	(0.01)	(0.50)
Number of dropout HH members	-0.06	0.24**	-0.07**	-0.00	4.34***
	(0.06)	(0.10)	(0.03)	(0.02)	(1.45)
Average Sampling Weight in Household	-0.08	-0.64***	0.23***	-0.01	-5.36***
	(0.08)	(0.09)	(0.03)	(0.01)	(1.36)
Constant	2.69***	0.40***	0.09***	0.02	11.47***
	(0.08)	(0.08)	(0.03)	(0.01)	(1.34)
Observations	2658	2618	2651	2581	2655
R-squared	0.010	0.061	0.036	0.020	0.051

Analysis performed within households, with spillover terms identified by households in which more than one girl was selected into the sample. Robust standard errors in parentheses, clustered at the EA level.

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

TABLE 6.

Analysis of Spillovers on Siblings who were Not a Part of the Study:

	Schooling rate among:			Never Married rate among:		
	All siblings	Brothers	Sisters	All siblings	Brothers	Sisters
VARIABLES	(1)	(2)	(3)	(4)	(5)	(5)
CCT Treatment	0.047	0.034	0.041	0.033	0.02	0.041
	(0.024)**	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)
UCT Treatment	-0.03	0.052	-0.09	-0.042	-0.008	-0.052
	(0.03)	(0.04)	(0.033)***	(0.03)	(0.03)	(0.028)*
Within-Village Control CCT	0.047	0.047	0.041	0.03	0.018	0.032
	(0.024)*	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Within-Village Control UCT	-0.001	0.033	-0.041	0.006	0.007	0
	(0.03)	(0.05)	(0.04)	(0.04)	(0.03)	(0.05)
Observations	2426	1930	1846	2426	1930	1846
R-squared	0.03	0.02	0.02	0.02	0.01	0.03

Analysis performed using information on all siblings who were not a part of the study, dependent variables represent average within these groups for core respondents who had siblings fitting each category. Robust standard errors in parentheses, clustered at the EA level.

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

TABLE 7

Analysis of Social Networks, Treatments Split

	Terms Enrolled	Cognitive Test Scores	Ever Pregnant	HIV Positive	Money spent by beneficiary on own consumption	Household Consumption Aggregate
	(1)	(2)	(3)	(4)	(5)	(6)
CCT	0.17***	0.20*	-0.02	0.02	1.93	5.05***
	(0.06)	(0.12)	(0.04)	(0.02)	(1.33)	(1.87)
UCT	0.09	0.21*	-0.06**	-0.02**	1.41	-0.54
	(0.08)	(0.12)	(0.03)	(0.01)	(1.32)	(1.42)
Within-Village Control CCT	0.01	0.19**	-0.01	0.00	0.94	2.71
	(0.06)	(0.08)	(0.03)	(0.01)	(1.14)	(1.71)
Within-Village Control UCT	-0.08	-0.09	-0.02	0.05	-0.16	-0.90
	(0.09)	(0.11)	(0.05)	(0.04)	(1.29)	(1.65)
Number of Treated Friends	-0.02	-0.10	0.03	-0.01	-0.31	-2.65**
for CCT Treatment Girls	(0.05)	(0.10)	(0.03)	(0.01)	(0.79)	(1.07)
Number of Treated Friends	-0.02	-0.13	-0.01	0.00	-0.50	0.82
for UCT Treatment Girls	(0.10)	(0.10)	(0.02)	(0.00)	(1.16)	(1.57)
Number of Treated Friends	0.15*	0.03	0.03	-0.01	-0.38	-4.96***
for CCT Untreated Girls	(0.08)	(0.10)	(0.05)	(0.01)	(0.87)	(1.22)
Number of Treated Friends	0.14*	-0.01	-0.05	-0.05	-1.46	2.61
for UCT Untreated Girls	(0.08)	(0.15)	(0.05)	(0.03)	(1.66)	(2.30)
Number of Treated Friends	0.02	-0.15	0.09	-0.01	-1.80	-1.73
for Pure Control Girls	(0.15)	(0.14)	(0.08)	(0.01)	(1.32)	(1.53)
Number of friends who are dropouts	-0.16***	-0.15***	0.12***	0.02	-0.59	-0.36
	(0.05)	(0.04)	(0.02)	(0.01)	(0.43)	(0.65)
Number of friends in same cluster	0.00	-0.09***	-0.00	-0.00**	-1.10***	-1.00***
	(0.02)	(0.02)	(0.01)	(0.00)	(0.20)	(0.35)
1 Matched Friend	-0.01	-0.14***	0.03	0.00	-0.46	-2.93**
	(0.04)	(0.05)	(0.02)	(0.01)	(0.74)	(1.19)
2 Matched Friends	0.00	0.02	0.04	-0.01**	-0.30	0.19
	(0.07)	(0.07)	(0.03)	(0.01)	(0.94)	(1.60)
3 Matched Friends	-0.03	-0.07	0.08*	-0.01	0.65	1.64
	(0.10)	(0.08)	(0.04)	(0.01)	(1.09)	(2.10)
4 Matched Friends	-0.10	-0.01	0.07	0.05	-0.86	-2.29
	(0.20)	(0.16)	(0.09)	(0.04)	(1.51)	(2.72)
5 Matched Friends	0.22	0.46**	-0.26***	0.01	-3.09	-1.96
	(0.14)	(0.21)	(0.07)	(0.02)	(2.16)	(5.42)
Constant	2.64***	0.22***	0.22***	0.03***	10.55***	24.29***
	(0.05)	(0.05)	(0.02)	(0.01)	(0.87)	(1.13)
Observations	2660	2620	2653	2583	2657	2663
R-squared	0.014	0.049	0.023	0.014	0.030	0.044

Analysis performed within social networks, as defined by the 'five closest friends' listed by core respondents at baseline. Robust standard errors in parentheses, clustered at the EA level.

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

TABLE 8.

		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 9.

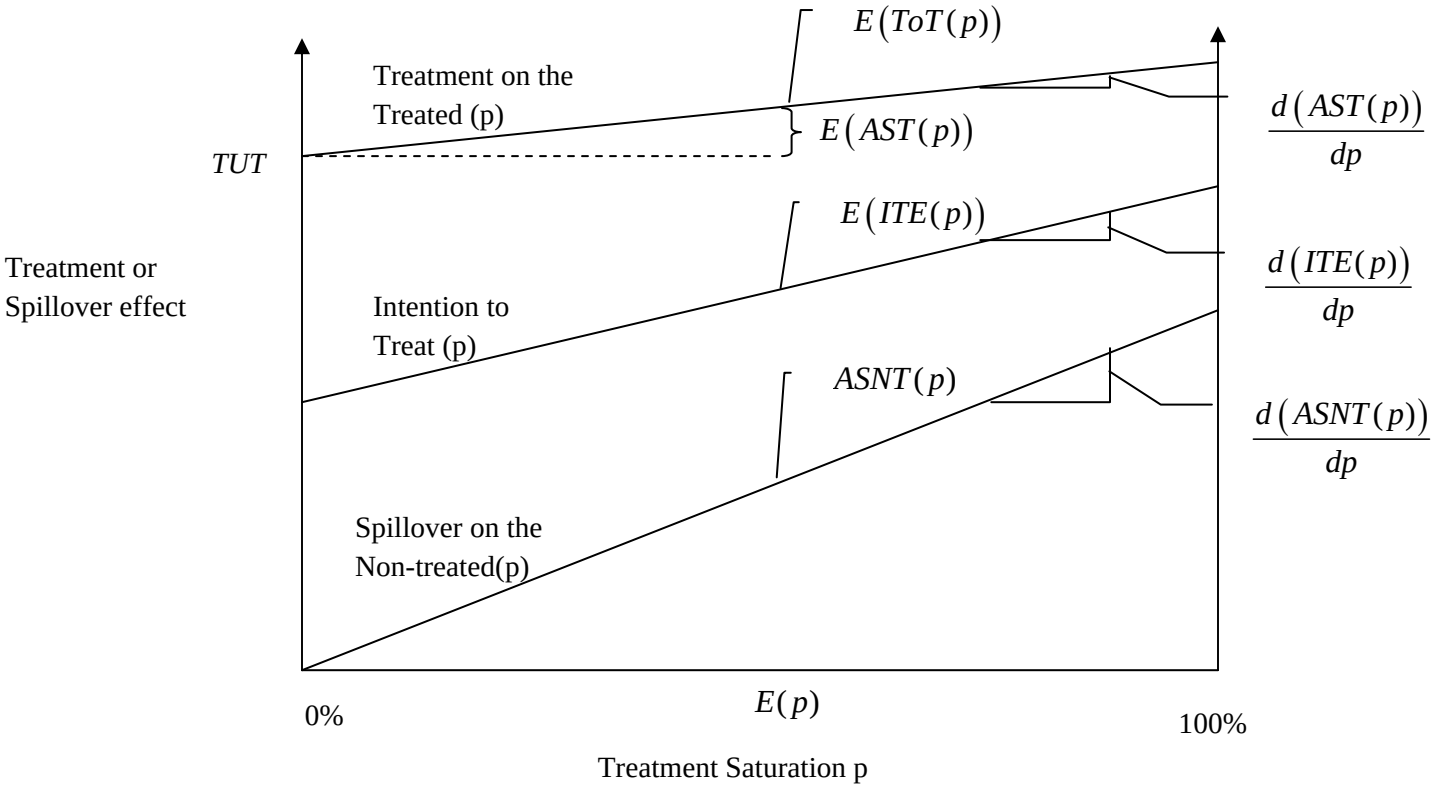
Robustness check using cross-EA variation in treatment intensity

	Dependent Variable:											
	Terms Enrolled		Cognitive Test Scores		Ever Pregnant		HIV Positive		Expenditures on girls' own consumption		Household Consumption Aggregate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CCT	0.119	0.119	0.136	0.048	0.022	-0.006	0.011	0.007	2.114	0.388	3.521	3.868
	(0.042)***	(0.083)	(0.049)***	(0.088)	(0.025)	(0.039)	(0.011)	(0.016)	(0.865)**	(1.157)	(1.170)***	(1.847)**
UCT	0.036	0.020	0.089	-0.040	-0.052	-0.079	-0.017	-0.018	0.518	-1.104	-1.305	-0.386
	(0.050)	(0.110)	(0.079)	(0.109)	(0.024)**	(0.048)	(0.008)**	(0.020)	(0.970)	(1.380)	(1.344)	(1.781)
Within CCT EA Control	0.012	0.015	0.181	0.191	0.001	0.000	0.000	0.001	0.268	0.285	0.989	0.966
	(0.047)	(0.047)	(0.068)***	(0.068)***	(0.024)	(0.024)	(0.009)	(0.009)	(0.774)	(0.788)	(1.328)	(1.366)
Within UCT EA Control	-0.100	-0.094	-0.198	-0.171	-0.007	-0.010	0.033	0.034	-1.407	-1.270	-1.356	-1.350
	(0.063)	(0.066)	(0.135)	(0.128)	(0.032)	(0.034)	(0.026)	(0.026)	(0.801)*	(0.833)	(1.366)	(1.381)
# of treated EAs within 3 km	-0.024	-0.022	-0.044	-0.023	0.007	0.004	0.003	0.004	-0.133	-0.089	-0.319	-0.383
	(0.018)	(0.020)	(0.025)*	(0.030)	(0.010)	(0.011)	(0.004)	(0.005)	(0.311)	(0.334)	(0.530)	(0.599)
# of treated EAs between 3 & 6 km	0.005	0.015	0.001	0.017	-0.001	-0.002	-0.002	-0.003	-0.838	-0.907	-0.810	-0.989
	(0.013)	(0.016)	(0.015)	(0.022)	(0.006)	(0.008)	(0.003)	(0.003)	(0.252)***	(0.275)***	(0.439)*	(0.545)*
# of total EAs within 3 km	0.015	0.013	0.042	0.030	-0.002	0.000	-0.001	-0.002	0.123	0.037	0.416	0.329
	(0.012)	(0.013)	(0.015)***	(0.018)	(0.006)	(0.007)	(0.003)	(0.003)	(0.206)	(0.231)	(0.299)	(0.326)
# of total EAs between 3 & 6 km	-0.002	-0.006	-0.007	-0.015	0.002	0.002	0.000	0.001	0.353	0.389	0.226	0.380
	(0.007)	(0.008)	(0.008)	(0.012)	(0.003)	(0.004)	(0.001)	(0.001)	(0.139)**	(0.137)***	(0.250)	(0.287)
Treated individual * # of treated EAs		0.002		-0.061		0.021		0.000		0.746		1.405
within 3 kilometers		(0.039)		(0.060)		(0.023)		(0.011)		(0.799)		(1.034)
Treated individual * # of treated EAs		-0.029		-0.051		0.005		0.004		0.197		0.443
between 3 and 6 kilometers		(0.024)		(0.032)		(0.013)		(0.005)		(0.399)		(0.735)
Treated individual * # of total EAs		0.004		0.039		-0.011		0.002		-0.165		-0.246
within 3 kilometers		(0.021)		(0.030)		(0.012)		(0.006)		(0.362)		(0.479)
Treated individual * # of total EAs		0.012		0.027		0.000		-0.003		-0.079		-0.456
between3 and 6 kilometers		(0.013)		(0.018)		(0.007)		(0.003)		(0.221)		(0.385)
Observations	2,579	2,579	2,612	2,612	2,650	2,650	2,572	2,572	2,649	2,649	2,646	2,646
R-squared	0.1	0.1	0.18	0.19	0.19	0.19	0.03	0.03	0.11	0.11	0.11	0.11

Notes: 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, and an indicator for ever had sex. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

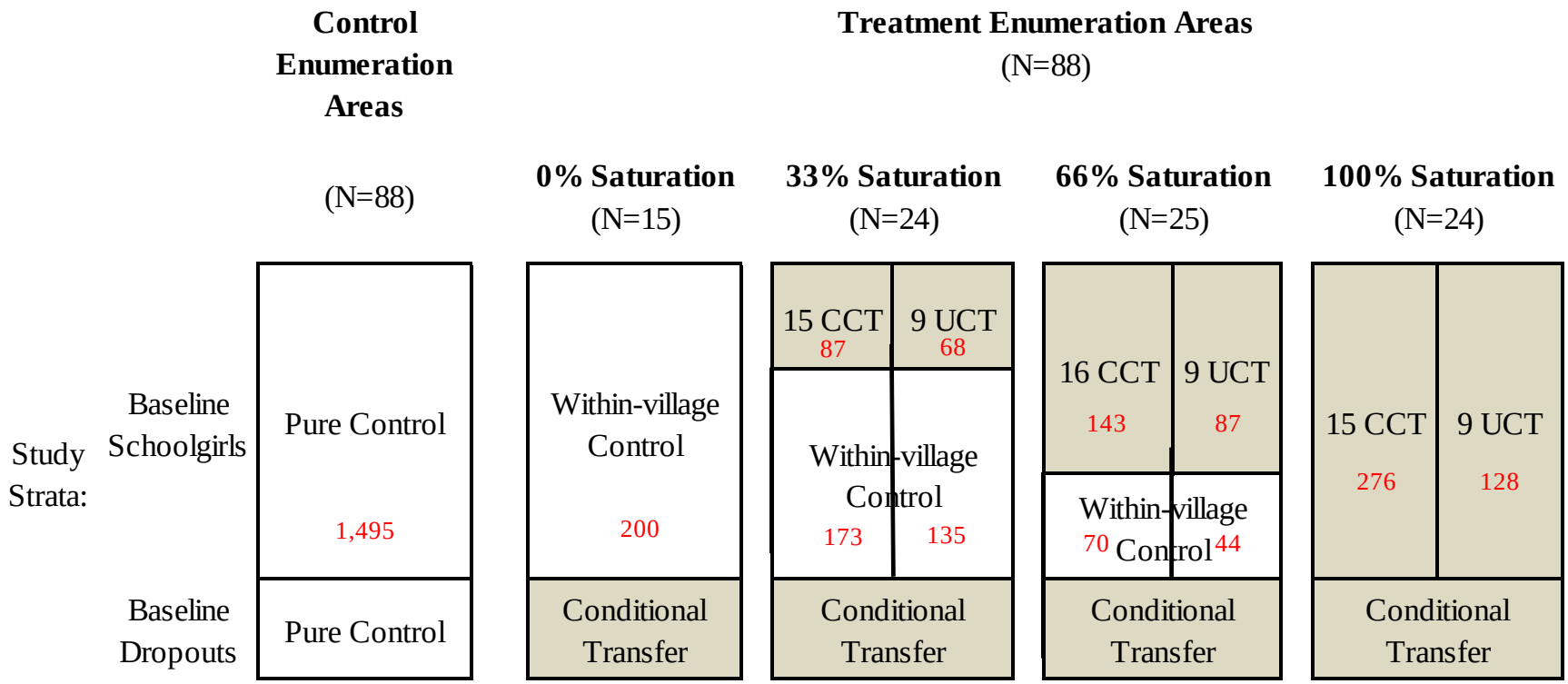
FIGURES.

Figure 1. Example of Spillover Effects.



Graphic demonstrates an example of a positive spillover effect that is stronger for untreated units than treated units. The compliance rate is .5 and we maintain Assumption 3, 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.

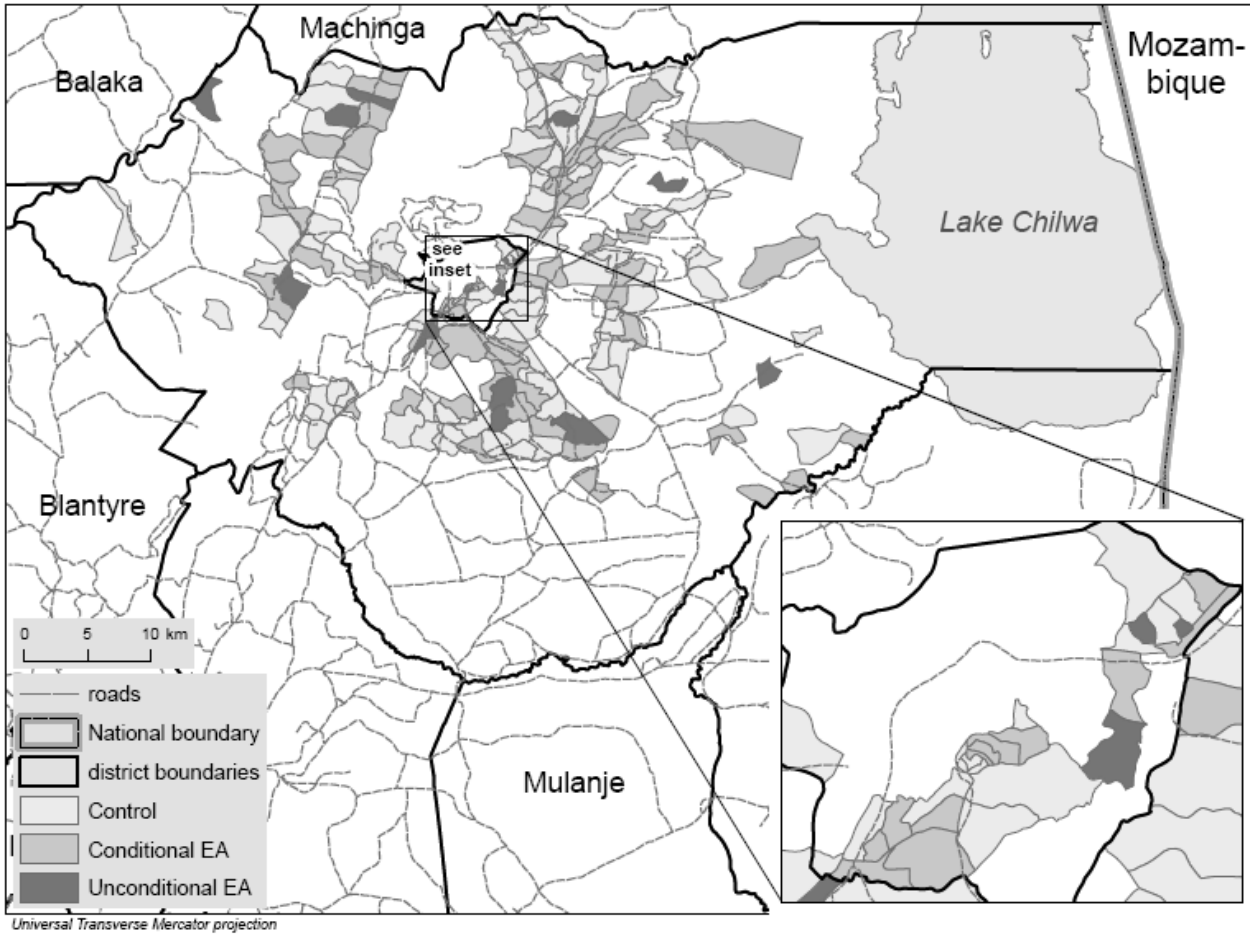


Figure 4. The Empirical Distribution of Saturations.

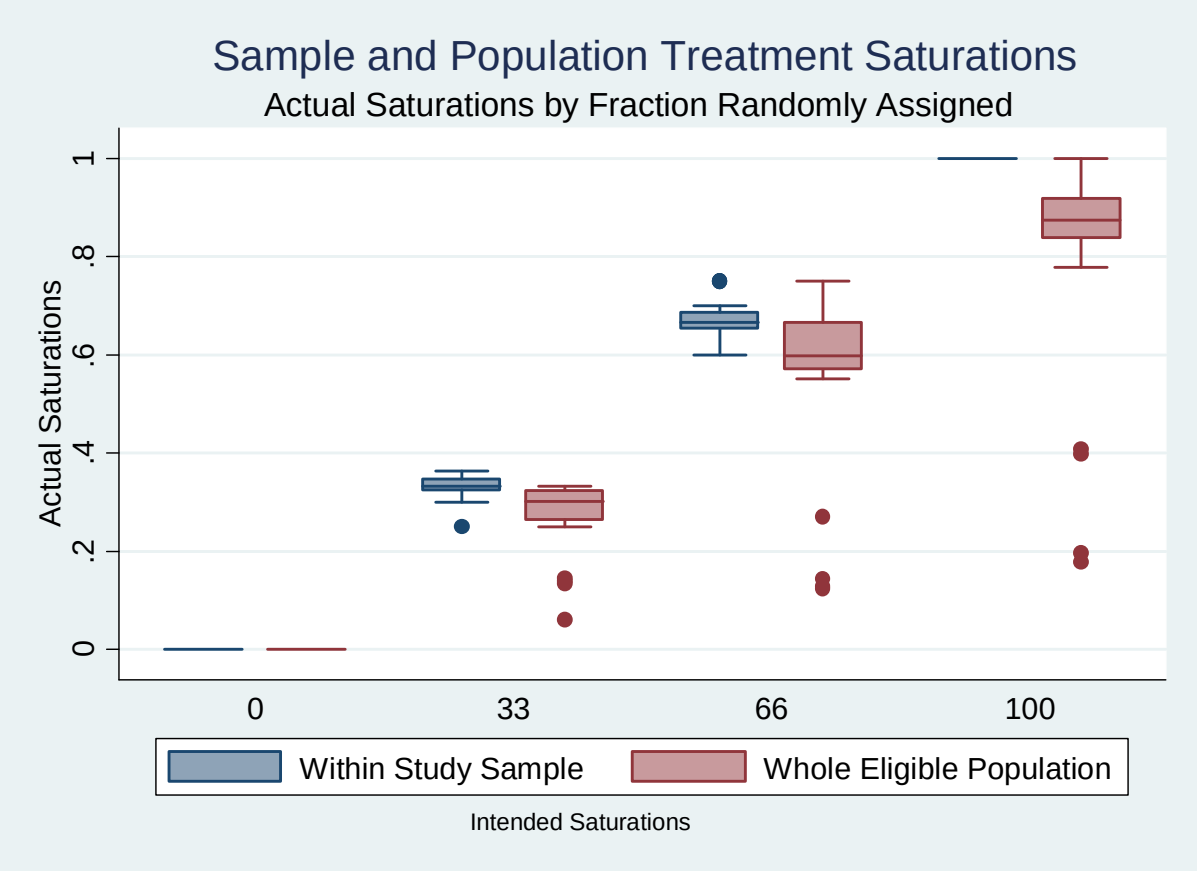


Figure 5. Saturation Impacts on Enrollment.

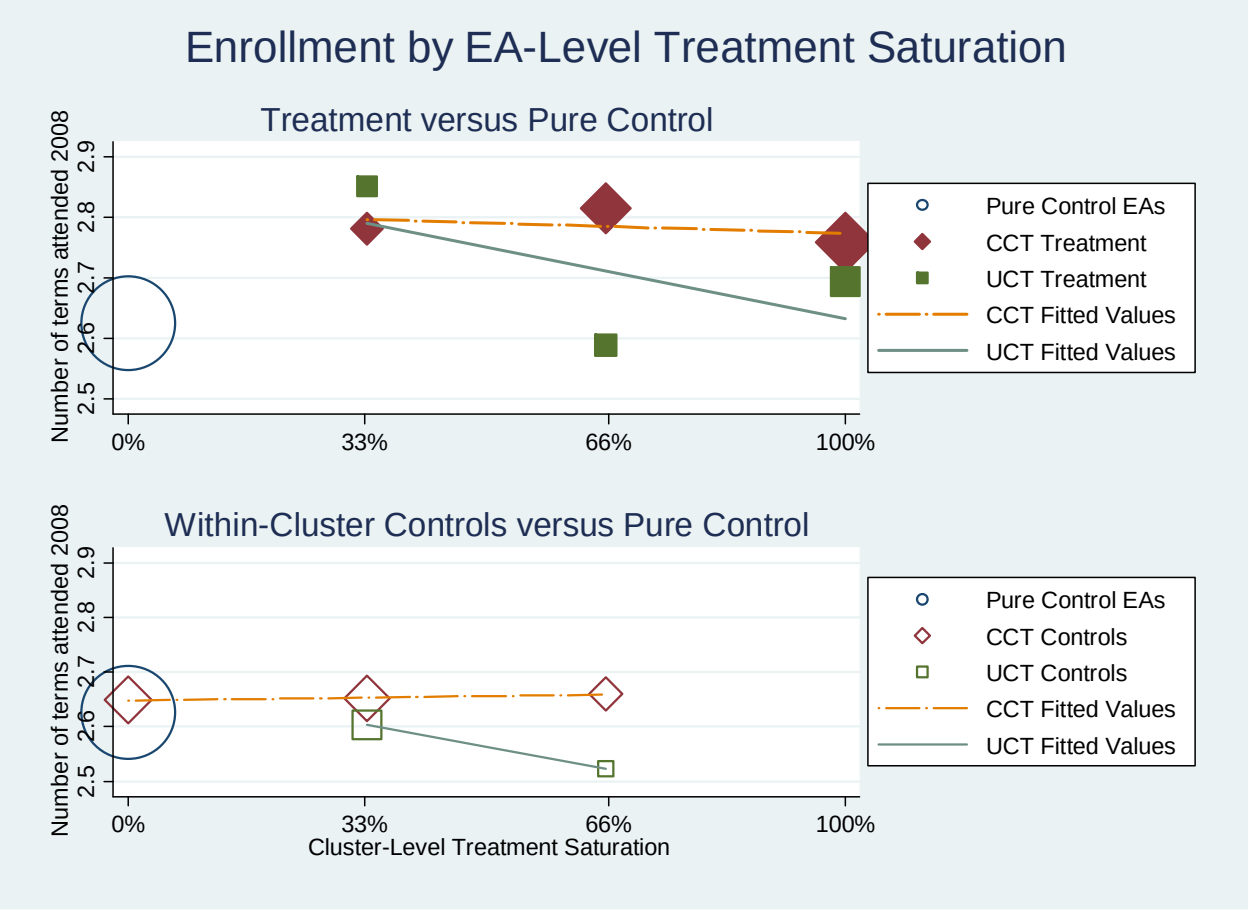


Figure 6. Saturation Impacts on Beneficiary Consumption in CCT EAs.

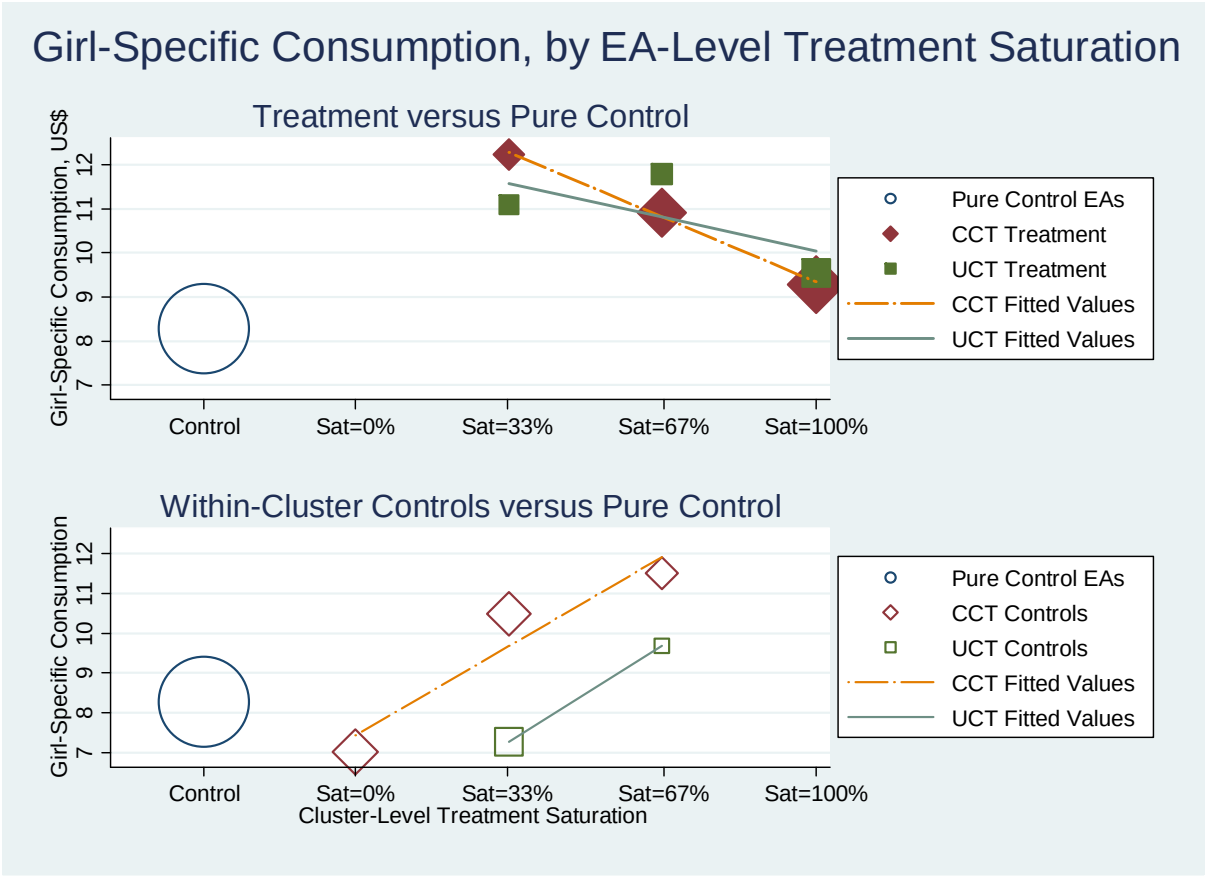


Figure 7. Comparison of Research Designs on Social Network Treatment Saturation.