Designing Experiments to Measure Spillover Effects*

Sarah Baird† J. Aislinn Bohren‡ Craig McIntosh§ Berk Özler¶

June 2015

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 randomly assigning individual treatment conditional on this cluster-level intensity, a novel set of treatment effects can be identified. We develop a formal framework for consistent estimation of these effects, and provide explicit expressions for power calculations. We show that the power to detect average treatment effects declines precisely with the quantity that identifies the novel treatment effects. A demonstration of the technique is provided using a cash transfer program in Malawi.

KEYWORDS: Experimental Design, Networks, Cash Transfers

JEL: C93, O22, I25

*We are grateful for the useful comments received from seminar participants at Caltech, Monash, Namur, Paris School of Economics, Stanford, University of British Columbia, UC Berkeley, University of Colorado, University of Melbourne, and Yale. We thank the Global Development Network, Bill & Melinda Gates Foundation, National Bureau of Economic Research Africa Project, World Bank's Research Support Budget, and several World Bank trust funds (Gender Action Plan, Knowledge for Change Program, and Spanish Impact Evaluation fund) for funding.

†George Washington University, sbaird@gwu.edu

‡University of Pennsylvania, abohren@sas.upenn.edu

§University of California, San Diego, ctmcintosh@ucsd.edu

¶World Bank, bozler@worldbank.org

1 Introduction

Multiple economic disciplines have begun to explore the empirical issues raised by spillover effects, or the impact an individual has on others. What Charles Manski (1993) refers to as endogenous effects are explored in different ways – by empirical studies permitting general equilibrium effects, the analysis of medical treatments that provide herd immunity, or by studies of network effects. An increasingly useful lens on this problem is experimental policy trials that explicitly consider interference between individuals. Once we permit interference in this context, the impact of a program solely on its beneficiaries becomes an unsatisfying measure of the real policy impact. Thus, it becomes more important to understand spillovers and the overall effect on the entire population. What if a program creates benefits to some only by diverting these benefits from others? How do treatment benefits depend on the intensity of treatment within a population? Does the study even have an unpolluted counterfactual?

These questions are of critical importance both for testing theoretical mechanisms and for designing optimal policy. Knowing the threshold at which herd immunity provides universal protection under an immunization campaign is important both for biomedical purposes and for cost effectiveness. Rapid reinfection of individuals treated for helminth worms or sexually transmitted infections may make health campaigns completely ineffective below some intensity of treatment and universally effective once a threshold has been cleared. Programs such as job training that alter the likelihood of receiving a benefit for which the size of the benefit pool is fixed may appear to confer large benefits upon the treated while having no total causal effect on the population at all. In many contexts a full understanding of the policy environment requires us to measure spillover and threshold effects that are not captured by (or even worse are sources of bias in) standard experimental designs.

The possibility of interference between individuals has traditionally been seen as the Achilles heel of randomized experiments. Standard experimental designs are unable to identify and measure spillovers: the blocked design produces biased estimates, while the clustered

design is not biased, but provides no information to estimate the extent of spillovers. Given these concerns, a new wave of empirical work has emerged in the past decade trying to relax the strong assumption of no interference, or that individuals are not affected by the treatment status of others. This literature includes studies that uncover network effects using experimental variation across treatment groups,[1] leave some members of a group untreated,[2] exploit plausibly exogenous variation in within-network treatments,[3] or intersect an experiment with pre-existing networks.[4]

One experimental method that allows for interference between individuals within the same exogenously defined cluster is a *partial population* experiment (Robert A. Moffitt 2001), in which clusters are assigned to treatment or control, and a subset of individuals are offered treatment within clusters assigned to treatment. This design partially overcomes the challenge of allowing for interference and yields valid estimates of treatment effects and spillover effects on untreated individuals. But it provides no exogenous variation in treatment saturation to estimate the extent to which program effects are driven by the intensity of treatment in a cluster.[5] Consequently, the most recent empirical approach has been to conduct a two-level randomization in which the share of individuals assigned to treatment within a cluster is directly randomized.[6]

[1] Matteo Bobba and Jeremie Gignoux (2013); Edward Miguel and Michael Kremer (2004).

[2] Manuela Angelucci and Giacomo De Giorgi (2009); Felipe Barrera-Osorio, Marianne Bertrand, Leigh Linden and Francisco Perez-Calle (2011); Gustavo J. Bobonis and Frederico Finan (2009); Esther Duflo and Emmanuel Saez (2003); Rafael Lalive and M. A. Cattaneo (2009).

[3] Philip S. Babcock and John L. Hartman (2010); Lori A. Beaman (2012); Timothy G. Conley and Christopher R. Udry (2010); Esther Duflo and Emmanuel Saez (2002); Kaivan Munshi (2003).

[4] Abhijit Banerjee, Arun G. Chandrasekhar, Esther Duflo and Matthew O. Jackson (2013); Jiehua Chen, Macartan Humphries and Vijay Modi (2010); Karen Macours and Renos Vakis (2008); Emily Oster and Rebecca Thornton (2012).

[5] Most extant partial population experiments feature cluster-level saturations that are either endogenous (Oportunidades) or fixed ((Duflo and Saez 2003), where they are typically set at 50%). PROGRESA/Oportunidades (Mexico) is perhaps the most-studied example of a partial population experiment. This program features a treatment decision at the cluster (village) level and an objective poverty eligibility threshold at the household level, so both eligible and ineligible individuals in treatment villages can be compared to their counterparts in the pure control group. PROGRESA has been used to examine spillover effects in several contexts (Jennifer Alix-Garcia, Craig McIntosh, Katharine R. E. Sims and Jarrod R. Welch 2013; Angelucci and De Giorgi 2009; Bobonis and Finan 2009). Other partial population experiments include Duflo and Saez (2003) and Peter Kuhn, Peter Kooreman, Adriaan Soetevent and Arie Kapteyn (2011).

[6] Abhijit Banerjee, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston and Nina Singh (2012); Matias Busso and Sebastian Galiani (2014); Bruno Crepon, Esther Duflo, Marc Gurgand, Roland Rathelot

In this paper, we formally define *randomized saturation* (RS) designs as a two-step procedure: (i) assign each cluster to a treatment saturation and (ii) assign each individual to a treatment status according to the assigned cluster saturation. An RS design is fully characterized by the set of possible saturations and the distribution used to map each cluster to a saturation. Next, we define a set of treatment and spillover estimands and present a set of sufficient assumptions for an RS design to consistently measure these effects. The first set of estimands measure the average direct treatment effect and indirect spillover effect on untreated individuals in treated clusters across the distribution of saturations (the intention to treat (ITT) and spillovers on the non-treated (SNT)), as well as the total causal effect (TCE), the average difference between outcomes in treated and control clusters. We then split the treatment effect into the isolated impact of treatment, the treatment on the uniquely treated (TUT) and the spillover effect on the treated (ST).[7] Average effects are identified both at specific saturations and pooled across all saturations. Next, we define a set of slope estimands that measure the change in treatment and spillover effects with respect to the treatment saturation.[8] Finally, we also show that one can recover an estimate of the treatment on the compliers effect (TOC) by assuming that the observed spillover effects on those not offered treatment are a reasonable proxy for the spillovers experienced by non-compliers. This technique is critical because interference between units within clusters violates the exclusion restriction in the standard technique of instrumenting for treatment with randomized assignment to identify the TOC. In the process, a RS design also allows the researcher to discover the extent to which observed correlations in outcomes within clusters were caused

and Philippe Zamora (2013); Xavier Gine and Ghazala Mansuri (2012); Betsy Sinclair, Margaret McConnell and Donald P. Green (2012).

[7]Betsy Sinclair (2014) suggests that the design of experiments with multiple saturations can be extended to identify complementarities in treatment effects, a similar concept to the ST we define.

[8]Graham (2010) Bryan S Graham, Guido W Imbens and Geert Ridder (2010) provide a set of estimands to measure social spillovers in a setting with two types. Their mean allocation response functions map the social composition of a group (fraction of each type) onto the expected outcome for each type and across types, and are similar in spirit to our ITT, SNT and TCE, viewing treatment as type. Similar to Graham (2010)Graham, Imbens and Ridder (2010), we assume that no spillovers occur across clusters to identify these estimands. They provide conditions for non-parametric identification and use these estimands to study the efficient reallocation of types.

by endogenous effects, thereby offering a solution to the reflection problem, albeit after the fact, and informing the design of future studies.

Next, we develop a technical framework to guide researchers in how to optimally design a RS experiment. All (non-trivial) RS experiments yield consistent estimates of treatment and spillover effects, but the power of each experiment varies with the saturation profile and the share of clusters assigned to each saturation. We impose a random effects variance structure and derive explicit expressions for the minimum detectable treatment and spillover effects, and minimum detectable effects to measure the shape of spillovers. These minimum detectable effects determine the statistical power of different RS designs.[9] Power is a function of some standard quantities, such as the effect size and the intra-cluster correlation (ICC) of outcomes, as well as of some unique features of the RS design, such as the share of individuals assigned to each treatment saturation and the variance of saturations. Power for the average ITT and SNT is decreasing in precisely the quantity that identifies the novel effects, namely, variation in the intensity of treatment across clusters. RS designs therefore generate a tradeoff compared with the standard blocked, clustered, and partial-population designs: while they allow the researcher to identify novel effects on both the treated and the non-treated, this comes at the cost of reduced power to detect average effects across all treated or within-cluster control individuals.

We use these power calculations to provide insight on design choices such as the optimal degree of variation in saturations and the size of the pure control. Sinclair, McConnell and Green (2012) and Sinclair (2014) also discuss the power implications of measuring spillover effects; it is straightforward to see that fixing a sample size, dividing it into more treatment groups (i.e. adding within-cluster controls) will lower power due to the reduced number of individuals in each group. However, a contribution of our paper is to show that there are additional power implications in the presence of intra-cluster correlation. There is also a

[9] Keisuke Hirano and Jinyong Hahn (2010) use the special case of the Manski (1993) linear-in-means model to provide similar power calculations for treatment and spillover effects in a partial population design with no intra-cluster correlation.

power loss for any estimate pooling individuals across saturations, due to the variance in treatment saturation across clusters.[10]

The remainder of the paper is structured as follows. Section 2 formally defines a randomized saturation design, outlines the assumptions required to use this design and defines novel estimands related to spillovers. Section 3 shows how to estimate these treatment and spillover effects, presents closed-form expressions for the power of these estimates and derives properties of the optimal randomized saturation design to detect different effects. Section 4 presents an application of the technique using a cash transfer experiment in Malawi, wherein the fraction of eligible school-aged girls offered treatment was randomized across clusters. Section 5 concludes. All proofs are in Appendix A. Appendix B presents additional possibilities that an RS design can identify. [11]

2 Identifying Treatment and Spillover Effects

In this section, we formally define a randomized saturation design, present a structural model of interference, and define treatment and spillover effects that a randomized saturation design identifies. We conclude with an example illustrating the different effects.

2.1 A Randomized Saturation Design

2.1.1 Background

One of the most basic design choices in any multi-level experiment is the question of allocating treatment to N individuals distributed across C clusters. Conventional wisdom focuses on the 'design effect', whereby a positive correlation between the outcomes of individuals in

[10] In the Supplemental Appendix, we provide a Matlab program that allows a researcher to calculate the power of different potential RS designs.

[11] First, we illustrate how an RS design can consistently estimate the pure control outcome, a useful result for situations in which institutional constraints prohibit including pure control outcomes. Second, we compute the power of an RS design to detect treatment effects when it is determined ex post that there are no spillover effects. Third, we show that an RS design can also be used to estimate spillover effects in overlapping networks, such as a friends or family network.

the same cluster, i.e. intra-cluster correlation (ICC), causes a power loss if treatment is assigned at the cluster level, and the *blocked* design, in which half of individuals in each cluster is treated and the other half is used as the counterfactual, emerges as optimal. Critically, however, individuals in the same cluster may behave similarly because they are influenced by the behavior of others in the group (endogenous effects), their behavior reflects the exogenous characteristics of the group (contextual effects), or because they share similar characteristics or face similar institutional environments (correlated effects) (Manski 1993). The entire thrust of the 'reflection problem' is the impossibility of separating these effects using observational data that is typically available at baseline. If only contextual or correlated effects are responsible for the observed ICC, indeed the blocked design proves optimal. However, if endogenous effects are present, then the counterfactual in a blocked design is contaminated by interference from treated individuals, and this design produces biased estimates of the treatment effect. A *clustered* design, in which some clusters are assigned to treatment and others to control, produces unbiased treatment effects if the endogenous effects occur within, but not across clusters. But this comes with the corresponding loss of statistical power arising from cross-cluster identification, and does not identify the extent of interference on either treated or untreated individuals. Thus this most basic of design choices ends up on the horns of the reflection problem. The RS design provides a solution to this conundrum.

2.1.2 The Design

A *randomized saturation* (RS) design is an experiment with two stages of randomization. Take as given a set of N individuals divided into C equal-sized, non-overlapping groups, or clusters.[12] The first stage randomizes the treatment saturation of each cluster, and the

[12]We assume clusters are equal in size to simplify the analysis. In practice, datasets may have significant variation in the size of the cluster and the researcher may want to group clusters into different sized bins. For example, rural and urban clusters. The RS design and the studies discussed here use a simple, spatially defined definition of 'cluster' that is mutually exclusive and exhaustive. This is distinct from the issue of randomizing saturations with overlapping social networks (Peter Aronow 2012), which typically require a more complex sequential randomization routine (Panos Toulis and Edward Kao 2013). However, an additional advantage of this design is that it will also create exogenous variation in the saturation of any network that is correlated with given cluster, even if this other network is overlapping. This is discussed in

second stage randomizes the treatment status of each individual in the cluster, according to the realized saturation of the cluster. Formally, in the first stage, each cluster $c = 1, ..., C$ is assigned a treatment saturation $\pi_c \in \Pi \subset [0, 1]$ according to the distribution F, where Π is the finite support of possible saturations.[13] Denote the mean saturation across clusters by μ and variance by η^2. In the second stage, each individual $i = 1, ..., n$ in cluster c is assigned a treatment status $T_{ic} \in \{0, 1\}$, where $T_{ic} = 1$ represents a treated individual.[14] The realized treatment saturation of stage 1 specifies the distribution of the treatment status in stage 2 for each cluster, $P(T_{ic} = 1|\pi_c = \pi) = \pi$. Let f be the probability mass function for distribution F. An RS design ω is completely characterized by the pair $\{\Pi, f\}$.

The saturation $\pi_c = 0$ represents a cluster with no treatment individuals, or a pure control cluster. A within-cluster control is defined as an untreated individual in a cluster with treated individuals: $S_{ic} = \mathbb{1}\{T_{ic} = 0, \pi_c > 0\}$. An RS design has the following distribution over the three possible treatment statuses:

$$\begin{aligned} \textbf{Treatment Individual:} \quad & P(T_{ic} = 1) = \mu \\ \textbf{Pure Control:} \quad & P(S_{ic} = 0, T_{ic} = 0) = f(0) := \psi \\ \textbf{Within-cluster Control:} \quad & P(S_{ic} = 1) = 1 - \mu - \psi := \mu_S \end{aligned}$$

We say a randomized saturation design has a *pure control* if $\psi > 0$.

The RS design nests several common experimental designs, including the clustered, blocked and partial population designs.[15] The blocked design is biased in the presence of spillovers, and it is not possible to measure spillovers with either design. Therefore, we

more depth in Appendix B.3.

[13] For expositional simplicity, we present the theoretical results in a discrete saturation support framework. In a finite cluster, $\Pi \subset \{0, 1/n, 2/n, ..., 1\}$.

[14] This notation implicitly assumes each cluster is of equal size. This is for notational convenience; the results easily extend to unequally sized clusters.

[15] Fixing the probability of treatment at P, the clustered design corresponds to $\Pi = \{0, 1\}$ and $f(1) = P$, the blocked design corresponds to $\Pi = \{P\}$ and $f(P) = 1$ and the partial population design corresponds to $\Pi = \{0, \pi\}$ and $f(\pi) = P/\pi$. In the clustered design, there is perfect correlation between the treatment status of two individuals in the same cluster and in the blocked design, there is no correlation. Note $\eta_T^2 = 0$ for all three.

must put some restrictions on the RS design in order to be able to identify treatment and spillover effects. We say a RS design is *non-trivial* if it has at least two saturations, at least one of which is strictly interior.

Definition 1 (Non-Trivial Design)**.** *A randomized saturation design is* non-trivial *if the support of* Π *contains at least* 2 *saturations and* $\exists\ \pi \in \Pi$ *such that* $\pi \in (0, 1)$*.*

Multiple saturations guarantee a comparison group to determine whether effects vary with treatment saturation, and an interior saturation guarantees the existence of within-cluster controls to identify spillovers on the untreated. Note that the blocked and clustered designs are trivial, while the partial population design is non-trivial.

An RS design introduces correlation between the treatment status of two individuals in the same cluster, which will affect the power of the design in the presence of intra-cluster correlation.[16] At one extreme, a clustered design introduces perfect correlation between the treatment status of individuals in the same cluster, while at the other extreme, individuals in a blocked design have no correlation in treatment status. As these two designs bracket the continuum of RS designs, it is natural that RS designs will introduce an intermediate level of correlation. This correlation is proportional to the variance of the cluster level treatment saturations, $\rho_T = \eta^2/(\mu(1-\mu))$, where η^2 can be split into the variance in treatment saturation across treated clusters, $\eta_T^2 = Var(\pi|\pi > 0)$, and the variance from pure control clusters:

$$\eta^2 = (1-\psi)\,\eta_T^2 + \left(\frac{\psi}{1-\psi}\right)\mu^2$$

This correlation and variance will play a key role in determining the power of an RS design.

2.1.3 Spillover Population and True Treatment Saturation

It is important to clarify the population in which the researcher is measuring spillovers. The RS design defines the treatment saturation of a cluster as the share of the study sample

[16] Note that the conditional correlation between the treatment status of two individuals, or the correlation conditional on these units being in the same cluster, is zero.

that is offered treatment. If spillovers only occur within the study sample, then this is the appropriate saturation measure.[17] Alternatively, if spillovers occur on individuals outside of the study sample, either because there is a 'gateway to treatment' and not all eligible individuals are sampled, or spillovers occur on a larger population than those eligible for treatment, then it is necessary to distinguish between the *true* treatment saturation (the share of treated individuals in the relevant spillover network) and the *assigned* treatment saturation (the share of treated individuals in the study population).[18] If the sampling rate and share of the spillover population eligible for treatment are constant across clusters, the true saturation is proportional to the assigned saturation. If sampling rates are driven by cluster characteristics or the share of the spillover population that is eligible for treatment differs across clusters, then the true saturation is endogenous. In this case, the researcher can instrument for the true saturation with the assigned saturation. To streamline the remainder of the theoretical analysis, we assume that the assigned and true saturations coincide.

2.2 The Structure of Interference

This section outlines three sufficient assumptions that jointly allow the use of an RS design for causal inference. The first two assumptions restrict how the treatment status of others impacts the outcome for an individual, and the third assumption is on the variance of the date generating process.

Let Y_{ic} represent the outcome for individual i in cluster c. In the most general framework, outcomes can depend in an arbitrary way on an individual's own treatment status, as well as

[17] For example, Banerjee et al. (2012) study interventions to improve performance among constables in Rajasthan police stations. Sinclair, McConnell and Green (2012) study sending social-pressure mailings to registered voters in a congressional district.

[18] For example, Gine and Mansuri (2012) sample every fourth household in a neighborhood, and randomly offer treatment to 80 percent of these households. This causes the true treatment saturation to be 20 percent rather than the assigned 80 percent. Other examples include unemployed individuals on official unemployment registries form a small portion all unemployed individuals in an administrative region (Crepon et al. 2013); neighborhoods eligible for infrastructure investments comprise only 3 percent of all neighborhoods (Craig McIntosh, Tito Alegria, Gerardo Ordonez and Rene Zenteno 2013); and malaria prevention efforts target vulnerable individuals, who account for a small share of total cluster population (GF Killeen, TA Smith, HM Ferguson, H Mshinda, S Abdulla et al. 2007).

the treatment status of all other individuals in the study. We express the *response function* for individual i in cluster c as

$$Y_{ic} = g_{ic}(\mathbf{T}; \varepsilon_{ic}) \tag{1}$$

where $\mathbf{T} \in \{0,1\}^{nC}$ is the treatment vector for all individuals and ε_{ic} is an individual error term.[19] The outcome Y_{ic} and treatment T_{ic} are observed, while the error term ε_{ic} is not. Let $(\mathcal{I}, \Omega, P)$ be a probability space over individuals $i \in \mathcal{I}$ and outcomes Ω. Our goal is to learn about the average effects on an individual drawn from $\mathcal{I}$.

The first assumption is the analogue of the stable unit treatment value assumption (SUTVA) for an RS design. We relax SUTVA within clusters, but maintain it across clusters: spillovers may flow within a cluster, but do not flow between clusters. This ensures that pure control clusters provide a valid counterfactual for treated clusters and that cross cluster comparisons can identify how spillovers vary with the intensity of treatment saturation. In the context of Equation 1, this means that Y_{ic} is independent of T_{jd} for all $d \neq c$ and $j = 1, ..., n$. Let $\mathbf{T_c} \in \{0,1\}^n$ be the treatment vector for cluster c.

Assumption 1 (SUTVA Across Clusters)**.** *For any two treatment vectors* $\{\mathbf{T_c}, \{\mathbf{T_d}\}_{d \neq c}\}$ *and* $\{\mathbf{T_c}, \{\mathbf{T_d}'\}_{d \neq c}\}$*,*

$$g_{ic}(\mathbf{T_c}, \{\mathbf{T_d}\}_{d \neq c}; \varepsilon_{ic}) = g_{ic}(\mathbf{T_c}, \{\mathbf{T_d}'\}_{d \neq c}; \varepsilon_{ic})$$

for all $i = 1, ..., n$ *and* $c = 1, ..., C$*.*

Assumption 1 lets us simplify the response function to

$$Y_{ic} = g_{ic}(\mathbf{T_c}; \varepsilon_{ic}).$$

This assumption alone is sufficient for computing consistent estimates of average treatment and spillover effects in a non-trivial RS design with a pure control. Such consistent estimates

[19] We present the theory without covariates for simplicity, but including covariates in empirical work does not change the insights presented here.

can be obtained without placing any structure on the interaction between individuals within the same cluster.

Two additional assumptions are necessary to characterize the standard errors of treatment and spillover effects and determine properties of the optimal experimental design. Within a cluster, we observe a single realization of the many potential configurations of individual treatment assignment at a given saturation. Our second assumption restricts how the identity of individuals receiving treatment impacts the outcome of other individuals in the same cluster.[20] We follow Eric J. Tchetgen and Tyler VanderWeele (2010) in using the 'Stratified Interference' assumption proposed by Michael Hudgens and Elizabeth Halloran (2008). This assumption says that the outcome of an individual is independent of the identity of the other individuals assigned to treatment, and is similar in spirit to the anonymous interactions assumption in Charles F. Manski (2013). Formally, given an individual i and treatment vector $\{T_{jc}\}_{j\neq i}$ in cluster c, let $\mathcal{P}(\{T_{jc}\}_{j\neq i})$ denote the set of permutations of treatment vector $\{T_{jc}\}_{j\neq i}$.

Assumption 2 (Stratified Interference)**.** *For any treatment vector* $\{T_{jc}\}_{j\neq i}$ *and permutations* $T, T' \in \mathcal{P}(\{T_{jc}\}_{j\neq i})$,

$$g_{ic}(T_{ic}, T; \varepsilon_{ic}) = g_{ic}(T_{ic}, T', ; \varepsilon_{ic})$$

for all $i = 1, ..., n$ *and* $c = 1, ..., C$.

This significantly simplifies the analysis and allows inference without possessing information about the underlying network structure within a cluster.[21]

Estimating the variance of treatment and spillover effects also requires an assumption on the variance of the data generating process. We assume a random effects error structure.[22]

[20] This is not an issue when there is no interference within clusters, as each unit has only two potential outcomes.

[21] In the absence of this assumption, a researcher would need to observe the complete network structure in each cluster, understand the heterogeneity in networks across clusters, and use a model of network-driven spillovers to simulate the variance in outcomes that could be generated by these networks.

[22] The random effects model assumes that v_c is uncorrelated with any covariates, which we have since treatment status is randomly assigned.

In an RS design, this error structure decomposes the clustering of outcomes into two components: (i) the extent to which outcomes are endogenously driven by treatment of others in the same cluster, and (ii) the statistical random effect in outcomes, which reduces the power of the clustered estimates but does not imply interference between units.

Assumption 3 (Random Effects Error)**.** *The data generating process has a random effects, additively separable error structure, with* $\varepsilon_{ic} = v_c + w_{ic}$*, common cluster component* $v_c \sim (0, \tau^2)$*, individual component* $w_{ic} \sim (0, \sigma^2)$ *and* (v_c, w_{ic}) *orthogonal to* (π_c, T_{ic}) *for all* $i = 1, ..., n$ *and* $c = 1, ..., C$*.*

This approach mirrors regression techniques typically used to analyze economic and medical experiments. It enables a direct comparison of the power of RS designs to the power of the canonical blocked and clustered designs, making explicit the impact that randomizing saturations has on power. This approach differs from the approach taken by the recent statistics literature (Hudgens and Halloran 2008), as well as in the paper most similar to ours (Sinclair, McConnell and Green 2012), both of which use randomization inference techniques (Ronald A. Fisher 1935).

Given Assumptions 1, 2 and 3, we can further simplify the response function to

$$Y_{ic} = g_{ic}(T_{ic}, \pi_c) + v_c + w_{ic}.$$

2.3 Defining Treatment and Spillover Effects

This section formally defines a set of treatment and spillover effects, both at specific saturations and pooled across multiple saturations. We focus on average effects across all individuals in the population.

2.3.1 Effects at Individual Saturations

The **Intention to Treat** (ITT) is the difference between the expected outcome for individuals offered treatment in a cluster with saturation π and the expected outcome for individuals

not offered treatment in a cluster with no treated individuals,

$$ITT(\pi) := E(Y_{ic} \mid T_{ic} = 1, \pi_c = \pi) - E(Y_{ic} \mid T_{ic} = 0, \pi_c = 0)$$

while the **Spillover on the Non-Treated** (SNT) is the analogous difference for individuals not offered treatment,

$$SNT(\pi) := E(Y_{ic} \mid T_{ic} = 0, \pi_c = \pi) - E(Y_{ic} \mid T_{ic} = 0, \pi_c = 0).$$

Individuals offered treatment will experience two types of treatment effects, a direct treatment effect from the program as well as a spillover effect that arises from the treatment of other individuals in their cluster. We formalize these effects by decomposing the ITT into two components, the **Treatment on the Uniquely Treated** (TUT),

$$TUT \quad := \quad E(Y_{ic} \mid T_{ic} = 1, \pi_c = \underline{\pi}) - E(Y_{ic} \mid T_{ic} = 0, \pi_c = 0),$$

and the **Spillover on the Treated** (ST),

$$ST(\pi) \quad := \quad E(Y_{ic} \mid T_{ic} = 1, \pi_c = \pi) - E(Y_{ic} \mid T_{ic} = 1, \pi_c = \underline{\pi})$$

where $\underline{\pi} := 1/n$ corresponds to a cluster with a single treated individual. The former measures the intention to treat on an individual, absent any spillover effects (equivalent to $ITT(\underline{\pi})$), while the latter measures the spillover effect on individuals offered treatment. The ITT is the sum of these two components, $ITT(\pi) = TUT + ST(\pi)$.

Definition 2 (Spillover Effects)**.** *There are* spillover effects *on the untreated (treated) if there exists a π such that $SNT(\pi) \neq 0$ ($ST(\pi) \neq 0$).*

Defining spillover effect on the intention to treat effects, rather than the treatment on the treated effects (formally defined in Section 2.3.3), ensures the definition includes both possible

channels for spillovers: (i) the effect of others' treatment on compliance and (ii) the effect of others' treatment on underlying outcomes. We can also use these definitions to determine what type of externality is generated by the treatment.

Definition 3 (Treatment Externalities)**.** *At saturation π, treatment creates a* negative externality *on the untreated (treated) if $SNT(\pi) < 0$ ($ST(\pi) < 0$) and a* positive externality *on the untreated (treated) if $SNT(\pi) > 0$ ($ST(\pi) > 0$). Treatment creates an* aggregate negative (positive) externality *if $\pi ST(\pi) + (1-\pi)SNT(\pi) < (>)\ 0$.*

The **Total Causal Effect** (TCE) measures the overall cluster-level effect on treated clusters compared to pure control clusters,

$$\begin{aligned} TCE(\pi): &= E(Y_{ic} \mid \pi_c = \pi) - E(Y_{ic} \mid \pi_c = 0) \\ &= \pi \times ITT(\pi) + (1-\pi) \times SNT(\pi). \end{aligned}$$

In the presence of spillovers, it is imperative to use the TCE, rather than the ITT, to inform policy, as the ITT may misrepresent the true effectiveness of the program. The TCE can be used to determine the most cost effective treatment saturation by comparing the total effect to the total cost of treating clusters at various saturations.

The TCE combined with the ITT can determine whether treatment effects are diversionary. We say that treatment effects are diversionary if the benefits to treated individuals are offset by negative externalities imposed on individuals in the same cluster.

Definition 4 (Diversionary Treatment Effects)**.** *Treatment effects are* partially (fully) diversionary *at saturation π if $TCE(\pi) > 0$ ($TCE(\pi) = 0$) and $SNT(\pi) < 0$.*

Diversionary treatment effects reallocate the distribution of value within a cluster to treated individuals, and the true effectiveness of the program is muted compared to the direct treatment effect.[23] In rare circumstances, treatment effects can even be more than fully diver-

[23]Of course, to say anything about the welfare implications of diversionary effects requires a welfare criterion specifying the social value of different distributions of the outcome variable within a cluster.

sionary. If the TCE is negative, the program causes aggregate harm even if treatment effects are positive.[24]

The **Value of Treatment** (VT) measures the individual value of receiving treatment at each saturation,

$$\begin{aligned} VT(\pi): &= E(Y_{ic} \mid T_{ic} = 1, \pi_c = \pi) - E(Y_{ic} \mid T_{ic} = 0, \pi_c = \pi) \\ &= ITT(\pi) - SNT(\pi). \end{aligned}$$

This effect can be used to determine whether the impact of treatment on the treated group differs from the impact of treatment on the untreated group. We can break value of treatment into two components, the direct benefit of treatment, TUT, and the difference in the spillover effect between treated and untreated individuals, $ST(\pi) - SNT(\pi)$, and use this to determine whether direct treatment effects and indirect spillover effects are complements or substitutes. If $VT(\pi)$ is decreasing in π, then the value of treatment is decreasing in the share of other individuals treated and spillover effects substitute for treatment, while if the VT is increasing in π, then the value of treatment is increasing in the share of other individuals treated and treatment is complementary with spillover effects. The VT yields important insight into the individual incentive to take up treatment. If the cost of undergoing treatment is independent of π, then as the treatment saturation rises, we will expect to see treatment uptake fall if treatment and spillover effects are substitutes, and rise if they are complements. Thus, treating at higher intensities will incentivize uptake in the case of complements, and it may be necessary to subsidize uptake at higher treatment intensities in the case of substitutes.

There is an intuitive relationship between the slope of the TCE, $VT(\pi)$ and the slope of the ITT and SNT. The slope of the TCE measures the aggregate marginal impact of adding

[24] We could also define a *diversionary spillover effect*, in which the spillover benefit to untreated individuals is offset by the negative impact on treated individuals in the same cluster.

another individual to treatment,

$$dTCE/d\pi \;\; = \;\; \pi\left(\frac{dITT}{d\pi}\right) + ITT(\pi) + (1-\pi)\left(\frac{dSNT}{d\pi}\right) - SNT(\pi).$$

Rearranging terms, this total marginal impact is equal to

$$dTCE/d\pi \;\; = \;\; VT(\pi) + \pi\left(\frac{dITT}{d\pi}\right) + (1-\pi)\left(\frac{dSNT}{d\pi}\right),$$

where the first term captures the individual benefit of adding another individual to treatment, and the remaining terms captures the marginal value to other individuals in the cluster of adding an additional individual to treatment. This expression yields insight into the trade-off faced by a policy maker in determining which saturation maximizes the TCE. Obviously, if treatment is directly beneficial and exhibits positive externalities, increasing the saturation will increase the TCE. Faced with marginal negative externalities, increasing the saturation will increase the TCE only if the marginal individual benefit exceeds the marginal negative externality.

2.3.2 Pooled Effects

It is possible to pool observations across clusters with different saturations and estimate an average, or pooled, effect for the entire experiment. Such pooled effects will be a weighted sum of the effect at each individual saturation. For RS design ω, we define the pooled ITT as the difference between the expected outcome for individuals offered treatment and the expected outcome for pure control individuals, giving *equal* weight to each saturation in the support of Π,

$$\begin{aligned}\overline{ITT}_\omega := \quad & \frac{1}{k}\sum_{\Pi\setminus\{0\}} E(Y_{ic} \mid T_{ic}=1, \pi_c=\pi) - E(Y_{ic} \mid T_{ic}=0, \pi_c=0) \\ = \quad & \frac{1}{k}\sum_{\Pi\setminus\{0\}} ITT(\pi),\end{aligned}$$

where $k := |\Pi| - 1$ is the number of non-zero saturations. The definitions for $\overline{SNT}_\omega$, $\overline{ST}_\omega$, $\overline{TCE}_\omega$ and $\overline{VT}_\omega$ are analogous, substituting $\pi_c = \underline{\pi}$ for $\pi_c = 0$ in $\overline{ST}$. The pooled measures can also be used to test for the presence of spillovers, externalities and diversionary effects. For example, a sufficient test for the presence of spillovers on untreated individuals is $\overline{SNT} \neq 0$.

Giving equal weight to each saturation in the RS design ensures that pooled effects are comparable across treatment and within-cluster controls, in that the pooled ITT, ST and SNT give the same weight to each saturation-specific effect. This also allows expression of the pooled VT as $\overline{VT} = \overline{ITT} - \overline{SNT}$. It is not possible to express the pooled TCE as a function of $\overline{ITT}$ and $\overline{SNT}$, since

$$\overline{TCE} = \frac{1}{k} \sum_{\Pi \setminus \{0\}} TCE(\pi) = \frac{1}{k} \sum_{\Pi \setminus \{0\}} \left(\pi ITT(\pi) + (1 - \pi) SNT(\pi) \right).$$

This is the case for any weighting scheme that gives the same weight to each saturation-specific effect for both the pooled ITT and SNT. In Section 3 we show how to weight the data to reflect the RS data structure.

2.3.3 Treatment on Compliers Effects

In a model with spillovers, the non-compliers in a treatment cluster may be affected by the treatment of compliers, and don't necessarily have the same expected outcome as non-compliers in a control cluster. Formally, the **Treatment on the Compliers** (TOC) effect is the difference between the expected outcome for individuals who comply with treatment and the expected outcome for pure control individuals who would have complied with treatment,

$$TOC(\pi) = E(Y_{ic} \mid T_{ic} = 1, R_{ic} = 1, \pi_c = \pi) - E(Y_{ic} \mid T_{ic} = 0, R_{ic} = 1, \pi_c = 0)$$

while the **Spillover on Non-Compliers** (SNC) effect is the difference between the expected outcome for individuals who do not comply with treatment and the expected outcome for

pure control individuals who would not have complied with treatment,

$$SNC(\pi) = E(Y_{ic} \mid T_{ic} = 1, R_{ic} = 0, \pi_c = \pi) - E(Y_{ic} \mid T_{ic} = 0, R_{ic} = 0, \pi_c = 0),$$

where $R_{ic} \in \{0, 1\}$ indicates whether an individual complies with treatment.[25] As is standard, R_{ic} is only observed for individuals with $T_{ic} = 1$, as the researcher does not know whether untreated individuals would have complied with treatment.

Combining these expressions, the $TOC(\pi)$ can be expressed as the difference between the $ITT(\pi)$ and $SNC(\pi)$, weighted by the compliance rate $r(\pi)$:[26]

$$TOC(\pi) = \frac{ITT(\pi) - (1 - r(\pi))SNC(\pi)}{r(\pi)}.$$

With no interference, $SNC(\pi) = 0$ and the standard approach of instrumenting for compliance with being offered the treatment identifies $TOC(\pi)$. With interference, we need an estimate of $SNC(\pi)$ to estimate $TOC(\pi)$. But $SNC(\pi)$ is not identified because compliance in the control is not observed, and therefore we must make an additional assumption to identify $TOC(\pi)$. A natural assumption is that the expected spillover on the non-compliers is equal to the spillover on the within cluster controls, a spillover effect that *is* identified in RS designs.

Assumption 4. *$SNC(\pi) = SNT(\pi)$ for all $\pi \in [0, 1]$.*[27]

Similar to the ITT, we can break the TOC into two effects: a direct treatment effect from the program, the **Treatment on the Unique Complier** (TUC), and a spillover

[25] The TOC is more commonly known as the Treatment on the Treated (TOT) effect. Throughout this paper, we use the term 'treated' to refer to the group offered treatment; therefore, to avoid confusion, we refer to the impact on those actually receiving treatment as the Treatment on the Compliers effect.

[26] If compliance varies across the saturation distribution, then changes in $ITT(\pi)$ will be driven by this as well as changes in the underlying $TOC(\pi)$ and $SNC(\pi)$. Indeed, in some cases, such as adoption of a new technology, the most important saturation-driven heterogeneity may come from variation in uptake across the saturation distribution.

[27] Crepon et al. (2013) estimate the treatment on the treated effect by assuming that the externality on an untreated worker is independent of his treatment status, which is equivalent to Assumption 4.

effect, the **Spillover on the Compliers** (SC), and define pooled effects, $\overline{TOC}$ and $\overline{SNC}$. If the compliance rate is constant with respect to treatment saturation, then an analogous expression exists for the pooled TOC, $\overline{TOC} = (\overline{ITT} - (1-r)\overline{SNC})/r$.[28]

2.4 Applied Examples of Spillovers

We illustrate the subtlety and importance of spillover and threshold effects with three stylized examples: measles vaccinations, deworming interventions and job training programs. First consider an intervention that vaccinates a share π of a cluster. The vaccination is almost fully protective independent of the treatment saturation, which means the $ITT(\pi)$ is flat with respect to π. However, the protection to the untreated only becomes sizeable as the saturation becomes high enough to provide herd immunity, and the $SNT(\pi)$ varies from zero to one. Thus, the $VT(\pi)$ is very large when vaccination rates are low and approaches zero at high vaccination rates since the unvaccinated are protected by herd immunity. Positive spillovers from treatment create a free-rider problem that may diminish the salience of vaccinations in populations that have very high overall treatment levels. This is illustrated in the left panel of Figure 1.

Deworming provides a more challenging case. Reinfection rates are proportional to the population prevalence of worm infections, which means that individuals who have received deworming treatment will quickly become reinfected in environments with high prevalence. The population treatment saturation drives long-term outcomes for both treated and untreated individuals, and effective deworming requires near universal treatment. The poignant irony of such a program is that the $VT(\pi)$ is close to zero at all saturations even though deworming can be effective if applied near universally. The key feature of this setting is the

[28] If the compliance rate varies with treatment saturation,

$$\overline{TOC} = \frac{1}{k} \sum_{\Pi \setminus \{0\}} \left[\left(\frac{1}{r(\pi)} \right) ITT(\pi) + \left(\frac{1 - r(\pi)}{r(\pi)} \right) SNC(\pi) \right].$$

and it is not possible to express $\overline{TOC}$ as a function of $\overline{ITT}$ and $\overline{SNC}$.

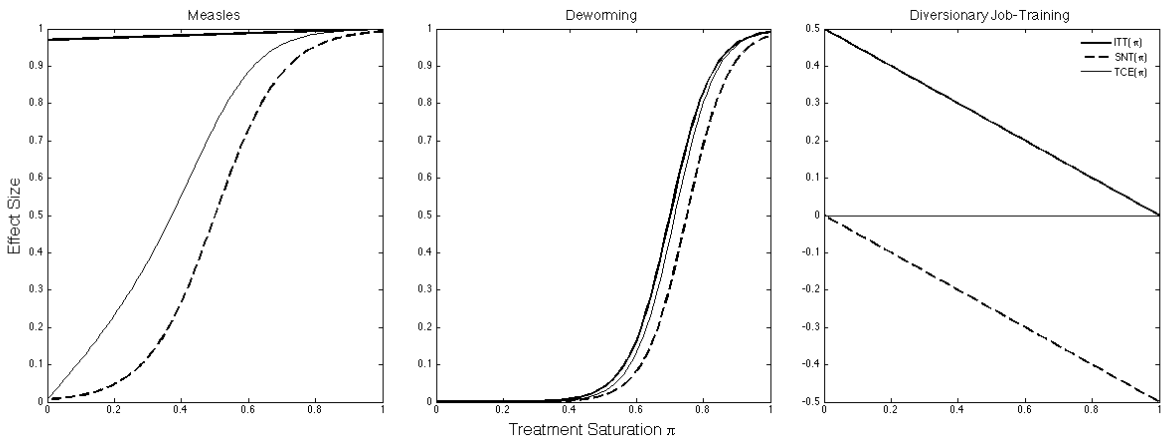

FIGURE 1. Examples

positive externality of treatment on both untreated and other treated individuals. This is illustrated in the center panel of Figure 1.

Another example is a job training program in which the training has no effect on the overall supply of jobs – treatment simply diverts benefits from untreated to treated individuals but provides little net benefit (Crepon et al. 2013). Similar examples are tutoring programs for admissions to college or grant-writing workshops that improve specific proposals for a fixed funding pool. This type of diversionary treatment effect will have a $TCE(\pi)$ that is zero for all π, even though the $ITT(\pi)$ and especially the $VT(\pi)$ are strictly positive. In the face of diversionary effects, a non-trivial RS design is imperative to identify the total policy effect, which is zero. Using within-cluster controls as counterfactuals will yield mistaken conclusions that the overall impact of a program is positive. This is illustrated in the right panel of Figure 1.

3 Minimum Detectable Effects and Optimal RS Designs

In order to conduct hypothesis tests and construct confidence intervals, we need to estimate the standard errors of the treatment and spillover effects. First, we examine how to estimate pooled effects in a simple linear regression model, which yields the most powerful test for the presence of treatment and spillover effects, and derive properties of the optimal RS design to detect pooled effects. Second, we address the question of how to estimate spillover effects at individual saturations and draw inference about the shape of spillover effects. We present non-parametric and parametric regression models, and derive properties of the optimal design to detect linear slope effects. Finally, we illustrate the power trade-off between measuring pooled and linear slope effects. We provide Matlab code for power calculations to help applied researchers evaluate the tradeoffs inherent in RS designs.

3.1 Pooled Effects

The pooled estimates provide the most powerful tests for the presence of treatment and spillover effects and whether treatment creates a positive or negative externality on untreated individuals. If it is not possible to detect a pooled effect, then it will not be possible to detect an effect at any individual saturation.

3.1.1 Estimating Pooled Effects

A simple difference-in-difference regression model to estimate pooled effects is

$$Y_{ic} = \beta_0 + \beta_1 T_{ic} + \beta_2 S_{ic} + \varepsilon_{ic}. \tag{2}$$

In order to estimate the $\overline{ITT}$ and $\overline{SNT}$ we define in Section 2.3, we must estimate Equation 2 with weights. When individuals are pooled across saturations, this unintentionally places

a disproportionate weight on treated individuals in high saturation clusters and untreated individuals in low saturation clusters. Similar to sampling weights, what we call saturation weights correct for the different probability of being assigned to treatment or within-cluster-control in each saturation bin, and the different probability of being assigned to different saturation bins.[29]

Definition 5 (Saturation Weights)**.** *Saturation weights apply weight* $s_\pi^T = 1/(\pi f(\pi))$ *to treated individuals and weight* $s_\pi^U = 1/((1-\pi)f(\pi))$ *to untreated individuals in treated clusters.*

For any non-trivial RS design with a pure control, estimating Equation 2 with saturation weights identifies $\widehat{\overline{ITT}}_\omega = \hat{\beta}_1$, $\widehat{\overline{SNT}}_\omega = \hat{\beta}_2$ and $\widehat{\overline{VT}}_\omega = \hat{\beta}_1 - \hat{\beta}_2$. Given Assumption 4 and assuming the compliance rate is constant with respect to π, Equation 2 identifies $\overline{TOC} = ((\hat{\beta}_1 - (1-\hat{r})\hat{\beta}_2))/\hat{r}$ where $\hat{r}$ is a consistent estimate of the compliance rate.[30] Sufficient tests for the presence of spillover effects are $\hat{\beta}_1 \neq 0$ and $\hat{\beta}_2 \neq 0$, while a one-tailed test of the sign of $\hat{\beta}_2$ determines whether treatment creates a negative or positive externality on untreated individuals and $\hat{\beta}_1 \neq \hat{\beta}_2$ determines whether there is direct value to treatment.

Different saturation weights are necessary to estimate the $\overline{TCE}$ because, by definition, this effect weights treated individuals proportional to π and untreated individuals proportional to $1-\pi$, in order to measure the overall effect on a cluster assigned to saturation π. Estimating Equation 2 with saturation weights $s_\pi = 1/f(\pi)$ returns $\hat{\beta}_1'$, which is an estimate of

$$\left(\frac{1}{\sum_{\Pi\setminus\{0\}}\pi}\right)\sum_{\Pi\setminus\{0\}}\pi ITT(\pi)$$

[29]For example, suppose an RS design assigns clusters to three saturations, $\Pi = \{0, 1/3, 2/3\}$ with equal probability, $f(\pi) = 1/3$ for each $\pi \in \Pi$. In a cluster assigned $\pi = 2/3$, an individual is twice as likely to be assigned to treatment as a cluster with $\pi = 1/3$. Weighting the treated individuals by $s_{2/3}^T = 3/2$ and $s_{1/3}^T = 3$ allows one to calculate the pooled estimate we define in Section 2.3, which places equal weight on both clusters, rather than twice as much weight on the $\pi = 2/3$ clusters.

[30]If the compliance rate varies with treatment saturation, we must weight observations to take into account the varying compliance rate in order to estimate $\overline{TOC}$.

and $\hat{\beta}_2'$, which is an estimate of

$$\left(\frac{1}{\sum_{\Pi\setminus\{0\}}(1-\pi)}\right)\sum_{\Pi\setminus\{0\}}(1-\pi)SNT(\pi).$$

Combining $\hat{\beta}_1'$ and $\hat{\beta}_2'$ yields an estimate of the pooled TCE,

$$\widehat{\overline{TCE}}_\omega = \left(\frac{1}{k}\right)\left(\hat{\beta}_1'\sum_{\Pi\setminus\{0\}}\pi + \hat{\beta}_2'\sum_{\Pi\setminus\{0\}}(1-\pi)\right).$$

Equation 2 doesn't identify $\overline{ST}$ or TUT.

3.1.2 Minimum Detectable Pooled Effects

Our first result characterizes the *Minimum Detectable Effect* (MDE) for the pooled effects estimated in Equation 2. The MDE is the smallest treatment or spillover effect that it is possible to distinguish from zero with a given power γ (Howard S. Bloom 1995). Suppose that the true effect is nonzero for some treatment or spillover effect β. Given statistical significance level α, the null hypothesis that $\beta = 0$ is rejected with probability γ (the power) for values of β that exceed:

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

We are interested in the relationship between the MDE, the data generating process and the choice of RS design. The MDE for the pooled effects depends on the size of the treatment and control group, and the within-cluster variation in treatment status, η_T^2. The following theorem characterizes the MDE of the pooled ITT and SNT.

Theorem 1 (Pooled MDE)**.** *Assume Assumptions 1, 2 and 3 and let ω be a non-trivial randomized saturation design with a pure control. The MDE of $\overline{ITT}_\omega$ for statistical significance*

level α and power γ is:

$$MDE_{\omega}^{T} = (t_{1-\gamma} + t_{\alpha})\sqrt{\frac{1}{nC}\left\{(n-1)\,\tau^2\left(\frac{1}{(1-\psi)\,\psi} + \left(\frac{1-\psi}{\mu^2}\right)\eta_T^2\right) + (\tau^2+\sigma^2)\left(\frac{\psi+\mu}{\mu\psi}\right)\right\}}$$

Substituting μ_S for μ yields an analogous expression for the MDE of $\overline{SNT}_{\omega}$, represented by MDE_{ω}^{S}.

This expression illustrates the relationship between the random effects structure and the RS design. The first term in the brackets captures the variation in $\widehat{\beta}$ due to the common cluster component of the error term, and the second term captures the variation in $\widehat{\beta}$ due to individual variation. Introducing randomization into the treatment saturation of clusters results in a power loss when there is a common cluster component to the error. Otherwise, if $\tau^2 = 0$, the standard error only depends on the size of the treatment and control groups, but is independent of how treatment is distributed across clusters.

3.1.3 Optimal Design to Detect Pooled Effects

Next, we derive the optimal RS design to test for the presence of treatment effects and spillover effects on the untreated. Consider the partial population design in which a cluster is treated with probability $1-\psi$, and treated clusters all have the same treatment saturation $P := \mu/(1-\psi)$. This design minimizes the variation in treatment saturation, and therefore, the MDE for treatment and spillover effects.

Corollary 1 (Optimality of Partial Population Design)**.** *Let Ω be the set of non-trivial RS designs with a pure control and suppose $\tau^2 > 0$. Then, fixing (μ, ψ) and letting $P = \mu/(1-\psi)$, a partial population design represented by*

$$\omega_{PP} = \{\Pi_{PP}, f_{PP}\} = \{\{0, P\}, \{\psi, 1-\psi\}\}$$

jointly minimizes MDE_{ω}^{T} and MDE_{ω}^{S} with respect to $\omega \in \Omega$.

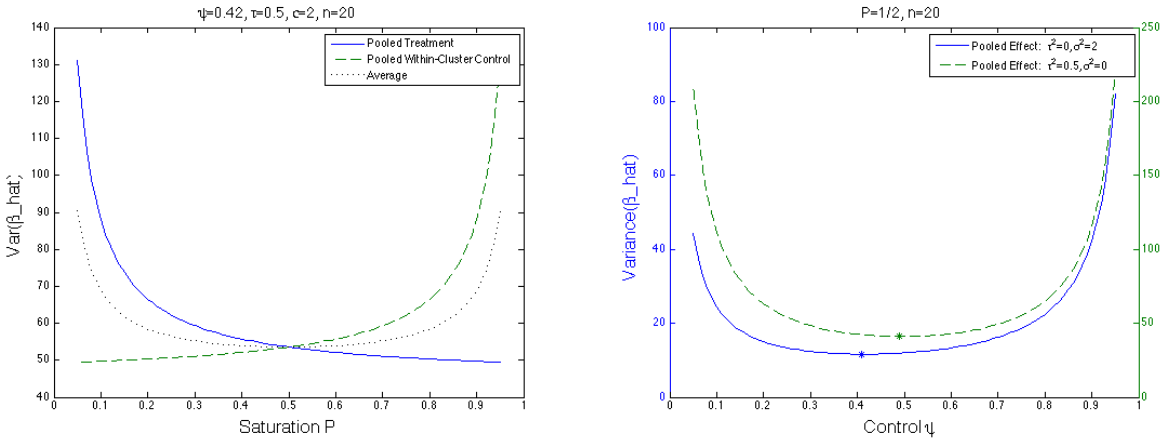

FIGURE 2. Partial Population Design

The optimality of a partial population design stems from a positive ICC.

Choosing the optimal treatment saturation P involves a trade-off. The power of the pooled ITT increases with P, while the power of the pooled SNT decreases with P. The relative importance of detecting these two effects, as well as their expected magnitudes, will determine the optimal P.

Corollary 2 (Optimal Saturation)**.** *Let Ω be the set of non-trivial RS designs with a pure control and suppose $\tau^2 > 0$. Then, fixing ψ, a partial population experiment with $P = 1/2$ minimizes $MDE^T_\omega + MDE^S_\omega$ with respect to $\omega \in \Omega$. In this design, $MDE^T_{PP} = MDE^S_{PP}$.*

The left panel of Figure 2 illustrates the variance of $\overline{ICC}$ and $\overline{SNT}$ in a partial population design, as a function of the saturation P.

The optimal size of the control group depends on the relative magnitude of the common cluster component of error to the individual component of error, and the size of the cluster.[31] A closed form expression for the optimal control is only possible at the extremes of $\tau = 0$ or $\sigma = 0$. At interior values of (τ, σ), we can show that the optimal control size lies in a

[31]For estimates of the parameter values relevant for power calculations, such as τ^2 and σ^2, one could use existing observational data or conduct a small pilot experiment. For example, Jinyong Hahn, Keisuke Hirano and Dean Karlan (2011) describe how to use data collected in the first stage of an experiment to inform the design of later stages, allowing for more efficient estimates of treatment effects. The reader should note, however, that the common cluster component of the error in observational data will include both endogenous effects and correlated effects, which would lead to an overestimation of τ^2 in the presence of spillover effects. A first-stage pilot experiment would not suffer from the same problem by separating these two effects.

relatively narrow range between approximately 0.41 and 0.5.[32]

Corollary 3 (Optimal Control Size)**.** *Let Ω be the set of non-trivial RS designs with a pure control. Then the size of the control that minimizes $MDE_\omega^T + MDE_\omega^S$ is*

$$\psi^*(\tau, \sigma, n) \in \left(\sqrt{2} - 1, \sqrt{n(1+n)} - n\right)$$

where $\sqrt{n(1+n)} - n < 1/2$ and $\lim_{n\to\infty} \sqrt{n(1+n)} - n = 1/2$.

1. *If $\tau = 0$, then $\psi^*(0, \sigma, n) = \sqrt{2} - 1$ for all (σ, n).*

2. *If $\sigma = 0$, then $\psi^*(\tau, 0, n) = \sqrt{n(1+n)} - n$ for all τ.*

Designating about 41% of individuals as pure controls yields the smallest sum of standard errors when there is no common cluster component to the error, while designating close to 50% is preferable when there is no individual component to error. It is always optimal to have the control be more than a third because it serves as the counterfactual for both treatment and spillover groups. As τ increases, the optimal number of control clusters increases. This comparative static arises because the variance in $\hat{\beta}$ due to individual error is proportional to the total number of *individuals* in each treatment group, while the variance in $\hat{\beta}$ due to correlated error is proportional to the total number of *clusters* in each treatment group. The right panel of Figure 2 illustrates the standard error of $\overline{ITT}$ and $\overline{SNT}$ in a partial population design with saturation $P = 1/2$ (note $\overline{ITT} = \overline{SNT}$ in this design), as a function of the control group size ψ, for two cases: (i) no common cluster component of error ($\tau = 0$) and (ii) no individual component of error ($\sigma = 0$). The minimum in each case is marked with an asterik.

Corollary 2 and Corollary 3 part (i) are similar in spirit to Hirano and Hahn (2010). They show how a partial population design can identify spillover effects in a linear-in-means model of with no ICC ($\tau = 0$). Their spillover effect of interest is the average spillover on

[32]It is straightforward to numerically compute the optimal control size for a specific (τ, σ, n).

all individuals in a treatment cluster (using our definitions, $\pi ST(\pi) + (1-\pi)SNT(\pi)$) and their treatment effect of interest is the average marginal impact of treatment, conditional on being in a treatment cluster ($ITT(\pi) - \pi ST(\pi) - (1-\pi)SNT(\pi)$). In this framework, they derive the optimal partial population saturation for a fixed size control group ($P = 1/2$), and then derive the optimal control group size ($\psi = \sqrt{2} - 1$).

Moving away from the partial population design to a design with variation in the treatment saturation, $\eta_T > 0$, leads to a power loss in the ability to measure pooled effects. This power loss increases more rapidly with respect to η_T^2 for settings with higher ICC, and designs with smaller control groups and treatment groups. Corollary 4 characterizes the rate at which this power loss occurs.

Corollary 4 (Linearity of Power Loss)**.** *Fix μ and ψ. Then $Var(\widehat{\beta})$ increases linearly with respect to η_T^2, with slope proportional to $\tau^2(1-\psi)/\mu^2$*

Taken together, these corollaries provide important insights on experimental design. If the researcher is only interested in detecting treatment effects and spillover effects on the untreated, then a partial population experiment has the smallest MDE, and Corollary 3 specifies the optimal control group size. However, partial population designs have the drawback that they only measure effects at a single saturation. When researchers care about the effects at multiple saturations, they will need to introduce variation in the treatment saturation. Corollary 4 establishes the rate at which the power of the pooled effects declines from this increase in treatment saturation variance.

3.2 Individual Saturation and Slope Effects

Now suppose that a researcher would like to determine how treatment and spillover effects vary with treatment intensity, or measure spillover effects on the treated. This section presents two methods to estimate these measures: (1) a non-parametric model that estimates an individual treatment and spillover effect at each non-zero saturation; and (2) a linearized

model that estimates the first order effect that changing the treatment saturation has on treatment and spillover effects. Identification of these models requires a RS design with multiple interior treatment saturations and a pure control. In addition to testing for the presence of treatment and spillover effects, it is possible to identify whether treatment creates a positive or negative externality on untreated and treated individuals and whether spillover effects are diversionary.

3.2.1 Non-Parametric Estimation

The most general regression model to determine the shape of spillover effects estimates the treatment and spillover effect at each saturation is

$$Y_{ic} = \beta_0 + \sum_{\Pi \setminus \{0\}} \beta_{1\pi} T_{ic} * \mathbb{1}\{\pi_c = \pi\} + \sum_{\Pi \setminus \{0\}} \beta_{2\pi} S_{ic} * \mathbb{1}\{\pi_c = \pi\} + \varepsilon_{ic}. \tag{3}$$

The support of the RS design determines which saturation specific estimates are identified, but unlike Equation 2, the definition of the coefficients is independent of the RS design. For any non-trivial RS design with multiple interior saturations and a pure control, Equation 3 identifies $\hat{ITT}(\pi) = \hat{\beta}_{1\pi}$, $\hat{SNT}(\pi) = \hat{\beta}_{2\pi}$, $\hat{TCE}(\pi) = \pi\hat{\beta}_{1\pi} + (1-\pi)\hat{\beta}_{2\pi}$ and $\hat{VT}(\pi) = \hat{\beta}_{1\pi} - \hat{\beta}_{2\pi}$ for each $\pi \in \Pi \setminus \{0\}$.[33] Given Assumption 4, the TOC is identified as:

$$\hat{TOC}(\pi) = (\hat{\beta}_{1\pi} - (1 - \hat{r}(\pi))\hat{\beta}_{2\pi})/\hat{r}(\pi),$$

where $\hat{r}(\pi)$ is a consistent estimate of the compliance rate at saturation π.

Tests for the presence of spillover effects at saturation π are $\hat{\beta}_{1\pi} \neq 0$ and $\hat{\beta}_{2\pi} \neq 0$. A one-tailed test of the sign of $\hat{\beta}_{2\pi}$ determines whether treatment creates a negative or positive externality on untreated individuals and $\hat{\beta}_{1\pi} \neq \hat{\beta}_{2\pi}$ determines whether there is direct value to treatment. It is also now possible to test for diversionary effects ($\pi\hat{\beta}_{1\pi} + (1-\pi)\hat{\beta}_{2\pi} \leq 0$ and $\hat{\beta}_{1\pi} \geq 0$) and the presence of spillover effects on treated individuals ($\hat{\beta}_{1\pi_j} - \hat{\beta}_{1\pi_k} \neq 0$).

[33] No saturation weights are necessary to estimate individual saturation effects.

Similarly, $\hat{\beta}_{2\pi_j} - \hat{\beta}_{2\pi_k} \neq 0$ tests whether $SNT(\pi)$ varies with the treatment saturation.

We can also use Equation 3 to estimate the change in spillover effects between saturations. Given saturations π_j and π_k, the rate of change of the spillover effect on treated individuals is

$$\delta^T_{jk} = \left(\beta_{1\pi_k} - \beta_{1\pi_j}\right) / \left(\pi_k - \pi_j\right),$$

with an analogous definition for the within-cluster controls, δ^S_{jk}. If spillover effects are affine, then this is a measure of the slope of the spillover effect, $dITT(\pi)/d\pi$ or $dST(\pi)/d\pi$. In the case of a non-linear spillover effect, it is a first order approximation of the slope. A one-tailed test of the sign of $\hat{\delta}^T_{jk} - \hat{\delta}^S_{jk}$ determines whether treatment and spillover effects are complements or substitutes.

3.2.2 Minimum Detectable Slope Effects

Our next main result characterizes the *Minimum Detectable Slope Effect* (MDSE) for the slope effects estimated in Equation 3. Similar to the MDE, the MDSE is the smallest rate of change in the effect between any pair of saturations $\pi_j, \pi_k \in \Pi$ that it is possible to distinguish from zero with a given power γ,

$$MDSE = \left[t_{1-\gamma} + t_\alpha\right] * SE\left(\hat{\delta}_{jk}\right).$$

The following theorem characterizes the MDSE of $ITT(\pi)$ and $SNT(\pi)$ between two saturations.[34]

Theorem 2 (MDSE)**.** *Assume Assumptions 1, 2 and 3 and let ω be a randomized saturation design with $\kappa \geq 2$ interior saturations. The MDSE of $ITT(\pi)$ between saturations π_j and π_k for statistical significance level α and power γ is:*

$$MDSE^T_\omega(\pi_j, \pi_k) = \frac{(t_{1-\gamma} + t_\alpha)}{\pi_k - \pi_j} \sqrt{\frac{1}{nC} * \left\{(n-1)\,\tau^2\left(\frac{1}{f(\pi_j)} + \frac{1}{f(\pi_k)}\right) + (\tau^2 + \sigma^2)\left(\frac{1}{\mu_j} + \frac{1}{\mu_k}\right)\right\}}$$

[34]The MDSE of $ITT(\pi)$ and $ST(\pi)$ are equivalent, by definition.

where $\mu_k := \pi_k f(\pi_k)$. *Substituting* $\mu_k^S := (1-\pi_k) f(\pi_k)$ *for* μ_k *yields an analogous expression for the MDSE of* $SNT(\pi)$*, denoted* $MDSE_\omega^S(\pi_j, \pi_k)$.

As the distance between two saturations increases, it is possible to detect smaller slope effects. At the same time, increasing the spread of saturations has a countervailing effect by making the number of treatment (within-cluster control) individuals very small at low (high) saturations. The latter effect dominates at saturations close to zero or one. When the cluster component of error is large, the share of clusters assigned to each saturation, $f(\pi_j)$, plays a larger role in determining the MDSE - a more equal distribution leads to a smaller MDSE. When the individual component of error is large, the share of treated and control individuals assigned to each saturation, μ_j, is more important. Note that while a pure control is required to identify treatment and spillover effects at each saturation in equation 3, it is not required to identify the slope effects.

An analogous result to Theorem 1 calculates the MDE of $ITT(\pi)$ and $SNT(\pi)$ for each saturation π; this result is presented in the Appendix A.3.

3.2.3 Optimal Design to Detect Slope Effects

There are two steps to the design choice for the non-parametric model: selecting which saturations to use (the support of Π), and deciding how to allocate individuals into each saturation bin (the distribution $f(\pi)$). A researcher can either fix a hypothesized slope size and determine how far apart saturations must be to detect this slope, or fix the distance between two saturations and calculate the smallest detectable slope size. Although a partial population design with a saturation of $\pi = 1/2$ is optimal for detecting pooled effects, this design does not identify slope effects. Moving away from the partial population design to a design with two interior saturations, Corollary 5 determines how we should assign the saturations.

Corollary 5 (Optimal Saturations)**.** *Let* Ω *be the set of RS designs with at least two interior*

saturations. Then, fixing $f(\pi_j) = f(\pi_k)$*, the saturations* (π_j^*, π_k^*) *that minimize*

$$MDSE_{\omega}^{T}(\pi_j, \pi_k) + MDSE_{\omega}^{S}(\pi_j, \pi_k)$$

are symmetric about $1/2$*. The optimal distance between saturations is*

$$\Delta^*(\tau, \sigma, n) \in \left(\sqrt{2}/2, 1\right).$$

Therefore, $\pi_j^* = (1 - \Delta^*)/2$ *and* $\pi_k^* = (1 + \Delta^*)/2$*.*

1. *If* $\tau = 0$*, then* $\Delta^*(0, \sigma, n) = \sqrt{2}/2$ *for all* (σ, n)*.*

2. $\Delta^*(\tau, \sigma, n)$ *is increasing in* τ^2 *and* n*, and decreasing in* σ^2*.*

3. *Asymptotically,* $\lim_{n\to\infty} \Delta^*(\tau, \sigma, n) = 1$*.*

Although Theorem 2 is generally too intractable to yield broader analytical insights about optimal design questions, it is possible to numerically calculate the MDSE for designs with more than two saturations. Given κ saturations, a researcher could use Theorem 2 to answer questions like (i) fixing equal sized bins $f(\pi_1) = ... = f(\pi_\kappa)$, what is the optimal spacing of saturations; or (ii) fixing equally spaced saturations $\pi_1, ..., \pi_\kappa$, what share of clusters should be assigned to each bin? This model also allows for hypothesis tests on the shape of the $ITT(\pi)$ and $SNT(\pi)$. For example, a test of concavity requires three interior saturations.

It is possible to use the expression for the $MDE(\pi)$ to calculate the optimal control group size numerically, given an estimate for τ^2 and σ^2.[35] Similar to the pooled model, the optimal size of the control group will be smaller in the presence of only individual error than in the presence of only cluster-level error, and will lie in between for intermediate error distributions. The optimal control will be smaller than the size of any treatment saturation $\psi^* < f(\pi)$, but will be larger than any treatment or within-cluster control group, $\psi^* > \max\{\pi f(\pi), (1 - \pi)f(\pi)\}$.

[35] The expression for $MDE(\pi)$ is in the proof of Theorem 2 in the Appendix.

3.2.4 Parametric Estimation

It is also possible to measure slope effects by imposing a functional form on the shape of the $ITT(\pi)$ and $SNT(\pi)$. For example, we could use an affine model to estimate the first order slope effect:

$$Y_{ic} = \delta_0 + \delta_1 T_{ic} + \delta_2 S_{ic} + \delta_3 (T_{ic} * \pi_c) + \delta_4 (S_{ic} * \pi_c) + \varepsilon_{ic} \tag{4}$$

This regression identifies the TUT as the intercept of the treatment effect, $\hat{TUT} = \hat{\delta}_1$. The coefficients δ_3 and δ_4 are slope terms estimating how effects change with the saturation, $d\hat{ST}(\pi)/d\pi = \hat{\delta}_3$ and $d\hat{SNT}(\pi)/d\pi = \hat{\delta}_4$. The intercept δ_2 estimates spillover effects at saturation zero. There should be no spillover effect on untreated individuals if the saturation of treatment is zero ($SNT(0) = 0$ by definition), so $\delta_2 = 0$ serves as a hypothesis test for the linearity of the spillover relationship. A test for $dST/d\pi = dSNT/d\pi$ is given by an F-test of the hypothesis that $\delta_3 = \delta_4$. Equation 4 also identifies $d\hat{SC}(\pi)/d\pi = (\hat{\delta}_3 - (1-r)\hat{\delta}_4)/\hat{r}$ and $\hat{TUC} = \hat{\delta}_1/\hat{r}$.

An analogous result to Theorem 2 characterizes the analytical expression for the MDSE in the affine model, which is proportional to $SE(\hat{\delta}_3)$ for treated individuals and $SE(\hat{\delta}_4)$ for untreated individuals.

Theorem 3 (Affine MDSE)**.** *Assume Assumptions 1, 2 and 3 and let ω be a randomized saturation design with $\kappa \geq 2$ interior saturations and a pure control. The MDSE of $ITT(\pi)$ for statistical significance level α and power γ is:*

$$MDSE^T_\omega = (t_{1-\gamma} + t_\alpha) \sqrt{\frac{1}{nC} * \{(n-1)\,\tau^2 h_1 + (\tau^2 + \sigma^2)\,h_2\}}$$

where

$$h_1 = \left(\frac{(\eta^2+\mu^2)^2 - 2\mu(\eta^2+\mu^2)E[\pi^3] + \mu^2 E[\pi^4]}{((\eta^2+\mu^2)^2 - \mu E[\pi^3])^2} \right) \text{ and } h_2 = \left(\frac{\eta^2+\mu^2}{(\eta^2+\mu^2)^2 - \mu E[\pi^3]} \right)$$

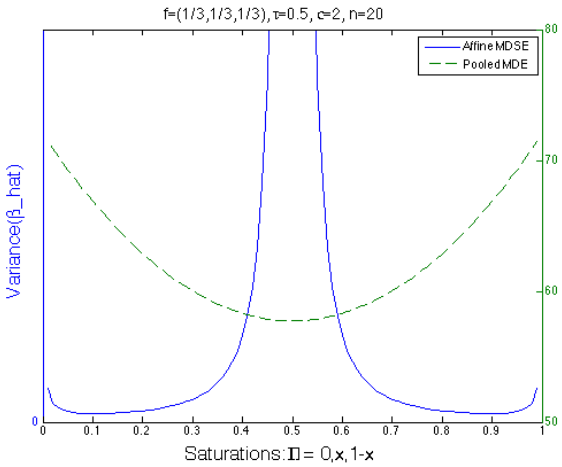

FIGURE 3. Trade-off between Pooled MDE and MDSE

An analogous expression characterizes the MDSE of $SNT(\pi)$*, denoted* $MDSE^S_\omega$*.*

A similar regression to Equation 4 can test for linearity or identify non-linear relationships. For example, including a squared term $T_{ic} * \pi_c^2$ would identify a quadratic relationship. In simulations, the affine $MDSE$ is smaller than the non-parametric $MDSE$ for detecting these higher moments. Another advantage of the affine model is that it can be estimated with data from a RS design in which saturations are assigned from a continuum.[36]

3.3 The Trade-off Between Estimating Pooled and Slope Effects

The optimal RS design for a slope analysis stands in sharp contrast to that for a pooled analysis, most obviously in the size of the pure control and the extent of variation in treatment saturation. A graphical representation of the tradeoff between detecting pooled and slope effects is presented in Figure 3. The optimal RS design to identify both slope and pooled effects will depend on the relative importance that the researcher places on each effect, as well as the expected size of each effect.[37]

[36] This design is necessary when using a Chow test to identify threshold effects.

[37] To facilitate actual implementation of an RS experiment, we created a Matlab program to calculate the minimum detectable effects in the pooled, non-parametric and affine models for different designs. The researcher specifies the relative importance of measuring (i) pooled versus slope effects and (ii) treatment versus spillover on the untreated effects. The program then calculates the optimal support of the RS design, Π, and the optimal allocation of clusters to each saturation bin, $f(\pi)$.

In conclusion, the RS framework provides an empirical resolution of why units within a cluster behave similarly. A study that finds high ICCs but no spillover effects can attribute clustering to correlated or contextual effects, while a study with the same ICCs but large spillovers should attribute clustering to endogenous effects. In this way the randomization of saturations resolves the reflection problem (albeit after the fact), and informs optimal design of subsequent experiments in similar contexts.

4 Empirical Application

We close with an empirical application of the technique using a cash transfer experiment in Malawi, wherein the fraction of eligible school-aged girls offered treatment was randomized across clusters.

4.1 Study Design

The Schooling, Income, and Health Risk (SIHR) study seeks to understand whether cash transfers could help adolescent girls improve schooling outcomes as well as delay marriage and pregnancy. In previous work, we have shown that, compared with a pure control group, conditional cash transfers (CCTs) significantly improved schooling outcomes while unconditional cash transfers (UCTs) caused substantial reductions in marriage and fertility rates among program beneficiaries (Sarah Baird, Craig McIntosh and Berk Özler 2011). However, spillovers are a potential concern for two distinct reasons. First, a large literature indicates that schooling cash transfer programs can alter the welfare of non-beneficiaries due to congestion effects in the classroom (Jere R Behrman, Piyali Sengupta and Petra Todd 2005), shifts in local norms around education (George A. Akerlof and Rachel E. Kranton 2002), income spillovers (Manuela Angelucci, Giacomo De Giorgi, Marcos A. Rangel and Imran Rasul 2010), or general equilibrium changes to prices (Jesse M. Cunha, Giacomo De Giorgi and Seema Jayachandran 2011) and production (Alix-Garcia et al. 2013). Second, and more

specific to the Malawian context, cash transfers can decrease young women's dependence on men for financial assistance (Winford Masanjala 2007) and/or the need for 'transactional sex' (Michelle J. Poulin 2007; Ann Swindler and Susan Watkins 2007), thereby reducing the incidence of teen pregnancies and early marriages among program beneficiaries, but with ambiguous spillovers to non-beneficiaries in the same communities. In this section, we exploit the sample of within-cluster controls to present the various treatment and spillover effects identified by RS designs.

The study took place in the Zomba district of Malawi. Before the start of the intervention, 176 EAs were selected from urban (Zomba city, 29 EAs) and rural (147 EAs) strata for inclusion in the study.[38] In the 176 study EAs, each dwelling was visited to take a census of all never-married females aged 13-22 years. Within this eligible population we defined two cohorts: those enrolled in school at baseline (baseline schoolgirls), and those not enrolled in school at baseline (baseline dropouts). All baseline dropouts were selected for inclusion in the study due to the small size of this cohort (approximately five per EA, accounting for about 15% of the target population), while we sampled within the larger cohort of baseline schoolgirls. The percentage of this cohort randomly selected for inclusion in the study was just above 60% and varied by geographical stratum and age group. [39] This sampling procedure yielded 3,796 individuals, who were enrolled in the study and completed baseline interviews at the end of 2007. Of these study participants, 889 were baseline dropouts and 2,907 were the baseline schoolgirls who we analyze here.

Out of the 176 EAs, 88 EAs were assigned to pure control and 88 to treatment. All baseline dropouts in treatment EAs were offered CCTs. The randomized saturation experiment as well as the UCT/CCT experiment was conducted only among baseline schoolgirls. 46 EAs had CCT saturations randomized, 27 EAs had the UCT saturations randomized,

[38]Each EA contains an average of 250 households spanning several contiguous villages. EAs were selected as the clusters for this study because they provide sampling frames with clearly delinieated official boundaries. Given the population and geographic size of an EA, it is plausible that SUTVA will hold between EAs. We test this assumption and present the findings in Table A2. A parallel analysis using village as the cluster yielded results that are not substantively different from the EA; we do not present these results in this paper.

[39]The sampling rate varied from 14% to 45% in urban EAs and 70% to 100% in rural ones.

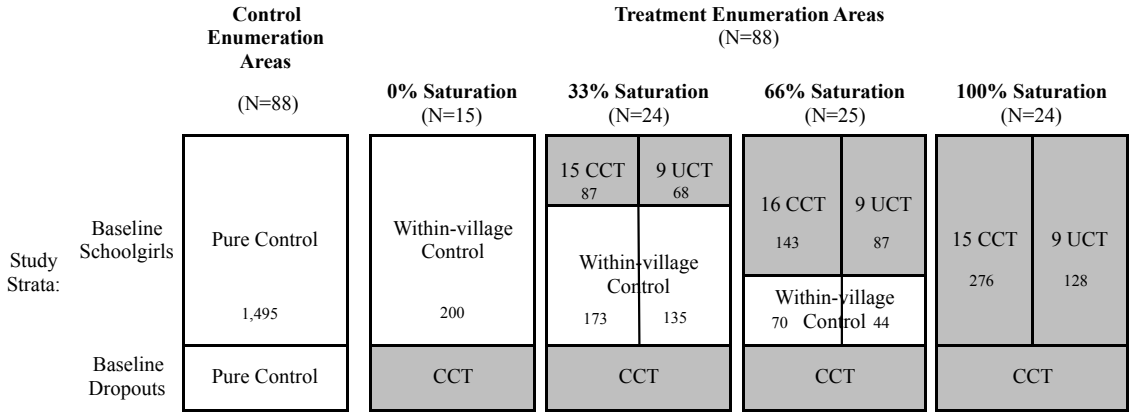

FIGURE 4. Research Design

Shaded cells indicate treatment and 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.

and 15 EAs saw only baseline dropouts treated. [40] In EAs assigned to CCT, 15 are treated at 33%, 16 are treated at 67%, and 15 are treated at 100%, while there were 9 UCT EAs in each saturation bin. The 15 EAs in which only baseline dropouts are treated provide a 0% CCT saturation, measuring the spillover from CCT treatment of baseline dropouts on baseline schoolgirls. Within each EA, we then selected the integer number of treatments that made the EA-level sample saturation as close as possible to that assigned. Figure 4 presents a schematic of the randomized saturation study design.

In the CCT arm, households were offered cash transfers of between $5 and $15 per month if the study participant attended school at least 80% of the days her school was in session during the past month. The UCT arm featured the same transfer system, but the cash transfers were offered unconditionally.[41] The cash transfer program started in early 2008, and continued for two years.

We performed the RS experiment only among baseline schoolgirls sampled into the study, meaning that the inclusion rules and sampling rates form a 'gateway to treatment' for the true saturation within both the eligible population and the overall population. A sampling rate

[40] Due to funding constraints for the transfers, the study included a larger pure control group than would have been ideal for power purposes alone.

[41] See Baird, McIntosh and Özler (2011) for more details on intervention design.

of less than 100% pushes down treatment saturations within the entire eligible population relative to those assigned within the study sample. Because this study featured a high overall sampling rate of 68%, the true saturations are only slightly lower than the assigned. The correlation coefficient between the assigned and true saturations at the EA level is 0.86, while the assigned saturations are completely orthogonal to the sampling weights with a correlation coefficient of 0.03. The actual saturation experiment is thus dampened by one-third from the assigned experiment. To recover marginal effects in the correct units of the true saturation, we instrument for the true saturations with the assigned.

Our analysis utilizes data from three sources. First, the annual SIHR Household Survey provides three rounds of data (baseline, 12-month follow-up, and 24-month follow-up). This survey provides data on the core respondent's marital status and fertility, as well as her network of friends. Second, we visited the schools of all study participants, who reported being enrolled in school during the 12-month follow-up interviews, and collected data on their enrollment and attendance directly from their schools.[42] Finally, to obtain an objective measure of learning, we administered independent tests for English, mathematics, and cognitive skills to study participants in their homes at the 24-month follow-up. 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.[43] The outcomes used in the empirical analysis, then, are enrollment, average test scores, and self-reported marriage and pregnancy.

4.2 Analysis of Treatment and Spillover Effects

Appendix Table A1 shows balance tests with the same specifications to be used in the analysis of spillover effects. All results are shown separately for CCTs and UCTs, providing

[42]While a school survey was also conducted at the 24-month follow-up, this was done only for a random sub-sample of study participants due to budget constraints. Hence, the outcome variable of number of terms enrolled goes from a minimum of zero to a maximum of three an indicator of school attendance during the first year of the program.

[43]Primary school in Malawi is from Standard 1 to 8, while secondary school is from Form 1 to 4.

cross-sectional baseline comparisons at the individual level while clustering standard errors at the EA level to account for the design effect. The set of 10 variables for which we examine baseline balance between various treatment groups is the same set reported in Baird, McIntosh and Özler (2011). Panel A shows the simple balance tests; the spillover sample is generally similar to the pure control at baseline. In Panel B we include linear slope terms, meaning that the top half of Panel B tests for the difference of the 0% saturation (observed in the CCT, extrapolated in the UCT) from the pure control. Hence, Panel B provides falsification for the intercept and slope terms to be used in the saturation analysis. Overall the experiment appears well balanced.

We now present the treatment and spillover effects that can be identified using the randomized saturation design. Table 1 estimates equations (2) and (4), with two modifications. First, we allow the CCT and UCT arms to have separate treatment and spillover effects. Second, we instrument for the true saturation within the eligible population using the randomly assigned saturation in the sample so as to provide marginal effects in the units of the true saturation. Our analysis includes all baseline schoolgirls, controlling for a basic set of baseline covariates and clustering standard errors at the EA level. We present two sets of results for each outcome, first showing simple ITT and SNT effects by estimating equation (2) in the odd-numbered columns, and then proceeding to test for the presence of saturation slope effects using equation (4) in the even-numbered columns. The bottom two panels of Table 1 explicitly calculate the treatment effects that were developed in Section 2.[44]

The regression coefficients on the treatment saturations give the linearized slope effects for each outcome. The pooled ITT is presented in rows 1 and 2 and the SNT in rows 3 and 4 of the odd-numbered columns. The TUT is the intercept term, given by the first two rows in the even-numbered columns. We can divide the TUT by the respective compliance rates to calculate the TUC, the treatment effect on the unique complier; and calculate the ToC, the pooled treatment on the compliers effect, using Assumption 4. These two estimands allow us

[44]The compliance rate was 77.4% for the CCT arm and 99% for the UCT arm.

to calculate the pooled spillovers on the compliers: $SC = TOC - TUC$. Finally, we perform F-tests on each of these estimands, which are linear combinations of regression coefficients across equations. Estimation conducted using Seemingly Unrelated Regressions with OLS models or two-step GMM with IV models provide identical results for the significance levels in the bottom panels of Table 1.

The cluster-level pooled spillover on the non-treated effects (SNT) are given by the coefficients on the within-cluster control indicators in the odd-numbered columns. Despite the sizable pooled intention to treat effects (ITT), we find no average spillover effects on the non-treated. Furthermore, for each statistically significant treatment effect the average spillover effect on within-cluster controls has the same sign, indicating no evidence of detrimental spillover effects among untreated individuals in treated clusters. The total causal effect, which is a weighted average of the ITT and SNT, presented in the bottom two panels confirms this finding: the TCE closely tracks the ITT in statistical significance and typically appears close to the ITT multiplied times .65, the average treatment saturation in clusters with any treatment. Saturation effects presented in the even-numbered columns suggest that these spillovers on the non-treated increase with treatment saturation, although none of these slope estimates are significant at the 10% level.

When we move to examining spillover effects on the treated (ST), we find some evidence that the beneficial treatment effects decline with treatment intensity. For example, the treatment on the uniquely treated (TUT) effect on enrollment in the CCT group is 0.25 terms, while the pooled ITT estimate is 0.133. Effects on test scores are evident in estimates presented in columns 3 and 4. Similarly, for marriage and pregnancy, the TUT effects in the UCT arm are consistently higher (in absolute value) than the pooled ITT effects, suggesting that beneficial intention to treat effects wear off as more eligible individuals are treated within the cluster. Intriguingly, this indicates that $\frac{d(SNT(\pi))}{d\pi}$ and $\frac{d(ST(\pi))}{d\pi}$ have opposite signs for all four outcome variables, suggesting a welfare tradeoff between treated and untreated units that becomes more pronounced as the treatment saturation increases.

Underlying the estimation of the saturation slope terms in Table 1 are the discrete distributions of the treatment saturations assigned in our experiment: 33%, 66%, 100%, and a 0% CCT cell that estimates the spillover on schoolgirls from CCT baseline dropout treatment alone. We calculate the non-parametric cell-specific $ITT(\pi)$ for each treatment saturation and $SNT(\pi)$ for each saturation below 100%. Table 2 presents this fully granular analysis of impact, showing coefficient estimates for each combination of treatment arm and saturation separately, using the non-parametric regression model in equation 3. In the first column we provide the average true saturation rate within the eligible population for each assigned saturation bin. The impact estimates in columns 2-5 reinforce the findings from Table 1, which used the affine model to estimate saturation effects: spillover effects on the non-treated are generally strongest (and have the same sign as the pooled intention to treat effects) for the cells with the highest treatment saturation (see, e.g. column 3 row 9 or column 5 row 11). Furthermore, again consistent with the earlier findings from Table 1, intention to treat effects are highest in the cells with low saturation, becoming insignificant for the highest saturations (see, e.g. the $ITT(\pi)$ estimates for schooling outcomes in the CCT arm and those for marriage and pregnancy in the UCT arm in the top panel of Table 2).

The purity of a cross-cluster counterfactual will be compromised if the regional intensity of treatment has an effect on outcomes. To test for this, we conclude by following Miguel and Kremer (2004) and Bobba and Gignoux (2013) in using GIS data on the locations of the EA centroids to count the number of treatment and control EAs within distance bands of <3km and 3-6km from each EA. Since the treated number of EAs is randomized conditional on the total number within each band, we can use this variation to look for cross-cluster spillovers that would violate Assumption 1. Table A2 demonstrates that this cash transfer experiment did not generate strong cross-cluster effects. Coefficient estimates for the number of treated EAs within the two distance bands are always small and statistically insignificant, implying that there are no spillovers for enrollment, test scores, marriage, or fertility across clusters

(columns 1, 3, 5, and 7).[45] Exploiting incidental randomization across clusters we confirm Assumption 1 and, as in the within-cluster analysis, find little evidence of spillover effects.

5 Conclusion

In recent years, empirical researchers have become increasingly concerned with the problem of interference between subjects. Experiments designed to rigorously estimate spillovers open up a fascinating set of research questions and provide policy-relevant information about program design. Research designs and RCTs that fail to account for spillovers can be biased; finding meaningful treatment effects but failing to observe deleterious spillovers can lead to misconstrued policy conclusions. This paper attempts to push the frontier of research designs by formalizing the analysis of randomized saturation experiments.

The benefit of randomizing treatment saturations is the ability to generate direct experimental evidence on the nature of spillover and threshold effects. The cost of doing so is statistical power. Having laid out the assumptions necessary to estimate both the mean and variance of spillover effects, we develop explicit, closed-form expressions for the power of RS experiments. We first provide a general expression for power when we seek to estimate treatment and spillover effects jointly. The power loss from randomizing saturations is directly related to the variation in treatment saturation, and so is an inherent feature of the design. Our explicit power calculation formulae provide concrete guidance for optimal research design depending on whether the researcher is primarily interested in measuring pooled treatment and spillover effects or slope effects (which necessitates more partially treated clusters). When spillover effects are found to be muted, this bolsters the credibility of causal inference from blocked designs.

Our empirical application provides little evidence of spillover effects within clusters, or

[45] In contrast to Bobba and Gignoux (2013), who find large spillover effects of PROGRESA in Mexico but only on treated individuals, we find no consistent evidence that program beneficiaries experience spillovers from adjacent clusters that are any different from untreated individuals (columns 1, 3, 5, and 7). In other words, the ST and the SNT measured cross-cluster are both zero.

indeed across clusters. This suggests that the significant decreases in marriage and fertility amongst schoolgirls in the unconditional cash transfer treatment group (Baird, McIntosh and Özler 2011) are causal in a larger sense, and are not arising because the treatment diverts such behavior to others girls in the study. For marriage, and pregnancy, the coefficient on treatment saturations for the within-cluster controls is in fact negative, indicating a slight protective effect of the program on nearby individuals who do not receive the treatment.

The framework presented here serves as an important guide to policy questions. For example, if a researcher is implementing a program with fixed resources and can either treat 100% of five villages or 50% of ten villages, which treatment allocation will maximize the total benefit? In the Malawi cash transfer program, our results suggest that they would have the same total effect, and the TCE of the program is closely approximated by the ITT times the average saturation rate, independent of how individuals are assigned to treatment. Small policy trials conducted on a subset of the population can miss important scale or congestion effects that will accompany the full-scale implementation of a program. To the extent that varying the cluster-level saturation leads to differential impacts on prices, norms, and congestion effects, the randomized saturation design provides an experimental framework that can bolster both external and internal validity.

References

Akerlof, George A., and Rachel E. Kranton. 2002. "Identity and Schooling: Some Lessons for the Economics of Education." *Journal of Economic Literature*, 40(4): 1167–1201.

Alix-Garcia, Jennifer, Craig McIntosh, Katharine R. E. Sims, and Jarrod R. Welch. 2013. "The Ecological Footprint of Poverty Alleviation: Evidence from Mexico's Oportunidades Program." *The Review of Economics and Statistics*, 95(2): 417–435.

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

Angelucci, Manuela, Giacomo De Giorgi, Marcos A. Rangel, and Imran Rasul. 2010. "Family networks and school enrolment: Evidence from a randomized social experiment." *Journal of Public Economics*, 94(3-4): 197–221.

Aronow, Peter. 2012. "A General Method for Detecting Interference in Randomized Experiments." *Sociological Methods Research*, 41(1): 3–16.

Babcock, Philip S., and John L. Hartman. 2010. "Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment." National Bureau of Economic Research, Inc NBER Working Papers 16581. NBER Working Papers.

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

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science*, 341(6144).

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. 2012. "Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police." National Bureau of Economic Research, Inc NBER Working Papers 17912.

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

Beaman, Lori A. 2012. "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S." *The Review of Economic Studies*, 79(1): 128–161.

Behrman, Jere R, Piyali Sengupta, and Petra Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change*, 54(1): 237–75.

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs." *Evaluation Review*, 19(5): 547–556.

Bobba, Matteo, and Jeremie Gignoux. 2013. "Policy Evaluation in the Presence of Spatial Externalities: Reassesing the Progresa Program." Working Paper.

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

Busso, Matias, and Sebastian Galiani. 2014. "The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment." National Bureau of Economic Research, Inc NBER Working Papers 20054.

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

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review*, 100(1): 35–69.

Crepon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment." *The Quarterly Journal of Economics*, 128(2): 531–580.

Cunha, Jesse M., Giacomo De Giorgi, and Seema Jayachandran. 2011. "The Price Effects of Cash Versus In-Kind Transfers." National Bureau of Economic Research, Inc NBER Working Papers 17456. NBER Working Papers.

Duflo, Esther, and Emmanuel Saez. 2002. "Participation and investment decisions in a retirement plan: the influence of colleagues' choices." *Journal of Public Economics*, 85(1): 121–148.

Duflo, Esther, and Emmanuel Saez. 2003. "The Role Of Information And Social Interactions In Retirement Plan Decisions: Evidence From A Randomized Experiment." *The Quarterly Journal of Economics*, 118(3): 815–842.

Fisher, Ronald A. 1935. *The Design of Experiments.* Oxford, England:Oliver & Boyd.

Gine, Xavier, and Ghazala Mansuri. 2012. "Together we will : experimental evidence on female voting behavior in Pakistan." Working Paper.

Graham, Bryan S, Guido W Imbens, and Geert Ridder. 2010. "Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach."

Hahn, Jinyong, Keisuke Hirano, and Dean Karlan. 2011. "Adaptive Experimental Design Using the Propensity Score." *Journal of Business & Economic Statistics*, 29(1): 96–108.

Hirano, Keisuke, and Jinyong Hahn. 2010. "Design of randomized experiments to measure social interaction effects." *Economics Letters*, 106(1): 51–53.

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

Killeen, GF, TA Smith, HM Ferguson, H Mshinda, S Abdulla, et al. 2007. "Preventing childhood malaria in Africa by protecting adults from mosquitoes with insecticide-treated nets." *PLoS Med*, 4(7): e229.

Kuhn, Peter, Peter Kooreman, Adriaan Soetevent, and Arie Kapteyn. 2011. "The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery." *American Economic Review*, 101(5): 2226–2247.

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

Macours, Karen, and Renos Vakis. 2008. "Changing Households' Investments and Aspirations through Social Interactions: Evidence from a Randomized Transfer Program in a Low-Income Country." World Bank Working Paper 5137.

Manski, Charles. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–542.

Manski, Charles F. 2013. "Identification of treatment response with social interactions." *The Econometrics Journal*, 16(1): S1–S23.

Masanjala, Winford. 2007. "The PovertyHIV/AIDS Nexus in Africa: A Livelihood Approach." *Social Science and Medicine*, 64(5): 1032–1041.

McIntosh, Craig, Tito Alegria, Gerardo Ordonez, and Rene Zenteno. 2013. "Infrastructure Impacts and Budgeting Spillovers: The Case of Mexico's Habitat Program." Working Paper.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217.

Moffitt, Robert A. 2001. "Policy Interventions, Low-Level Equilibria And Social Interactions." 45–82. MIT Press.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics*, 118(2): 549–599.

Oster, Emily, and Rebecca Thornton. 2012. "Determinants of Technology Adoption: Peer Effects in Menstrual Cup Take-Up." *Journal of the European Economic Association*, 10(6): 1263–1293.

Poulin, Michelle J. 2007. "Sex, money, and premarital relationships in southern Malawi." *Social Science and Medicine*, 65(11): 2383–2393.

Sinclair, Betsy. 2014. "Design and Analysis of Experiments in Multilevel Populations." In *Experimental Handbook, forthcoming.* , ed. Jamie Druckman, Donald P. Green and Skip Lupia.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science*, 56(4): 1055–1069.

Swindler, Ann, and Susan Watkins. 2007. "Ties of Dependence: AIDS and Transactional Sex in Rural Malawi." *Studies in Family Planning*, 38(3): 147–163.

Tchetgen, Eric J., and Tyler VanderWeele. 2010. "On Causal Inference in the Presence of Interference." *Statistical Methods in Medical Research*, 21(1): 55–75.

Toulis, Panos, and Edward Kao. 2013. "Estimation of Causal Peer Influence Effects." *Journal of Machine Learning Research*, 28. Proceedings of the 30th International Conference on Machine Learning Research.

TABLE 1. Linear Spillover Analysis

Dependent Variable:	Terms Enrolled		Average Test Score		Ever Married		Ever Pregnant	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CCT	0.133	0.250	0.026	0.054	-0.004	0.001	0.034	0.088
	(0.0454)***	(0.113)**	(0.00819)***	(0.0174)***	(0.021)	(0.074)	(0.028)	(0.071)
UCT	0.071	0.267	0.008	0.009	-0.069	-0.144	-0.056	-0.125
	(0.049)	(0.0933)***	(0.014)	(0.043)	(0.0247)**	(0.0482)***	(0.0229)**	(0.0563)**
Within CCT EA Control	0.018	-0.003	0.022	-0.008	0.009	0.007	0.008	-0.006
	(0.048)	(0.082)	(0.014)	(0.015)	(0.022)	(0.034)	(0.025)	(0.032)
Within UCT EA Control	-0.088	0.011	-0.015	-0.068	-0.004	0.036	-0.018	0.078
	(0.077)	(0.169)	(0.024)	(0.049)	(0.031)	(0.070)	(0.028)	(0.072)
True CCT treatment saturation, instrumented w/ assigned		-0.238		-0.061		-0.011		-0.113
		(0.227)		(0.0352)*		(0.128)		(0.141)
True UCT treatment saturation, instrumented w/ assigned		-0.386		-0.002		0.154		0.144
		(0.167)**		(0.076)		(0.115)		(0.110)
True CCT control saturation, instrumented w/ assigned		0.138		0.176		0.009		0.087
		(0.324)		(0.108)		(0.124)		(0.166)
True UCT control saturation, instrumented w/ assigned		-0.294		0.162		-0.123		-0.288
		(0.623)		(0.129)		(0.180)		(0.191)
Mean in Pure Control:	2.639		0.456		0.176		0.247	
Observations	2,579	2,579	2,612	2,612	2,649	2,649	2,650	2,650
Estimates for CCT: (average R3 compliance rate of 77.4%, compliance defined as attending school regularly)								
Intention to Treat (ITT)	0.133 ***		0.026 ***		-0.004		0.034	
Treatment on Uniquely Treated (TUT)	0.250 **		0.054 ***		0.001		0.088	
Spillovers on the Treated (ST)	-0.117		-0.028 *		-0.005		-0.054	
Treatment on Compliers (ToC)	0.167 ***		0.027 **		-0.008		0.042	
Treatment on Unique Complier (TUC)	0.323 **		0.070 ***		0.001		0.113	
Spillover on Compliers (SC)	-0.156		-0.043 **		-0.009		-0.072	
Spillover on Non-Treated (SNT)	0.018		0.022		0.009		0.008	
Total Causal Effect (TCE)	0.082 **		0.021 ***		-0.002		0.016	
Estimates for UCT: (average R3 compliance rate of 99%, compliance defined as receiving transfer)								
Intention to Treat (ITT)	0.071		0.008		-0.069 ***		-0.056 **	
Treatment on Uniquely Treated (TUT)	0.267 ***		0.009		-0.144 ***		-0.125 **	
Spillovers on the Treated (ST)	-0.196 **		-0.001		0.075		0.069	
Treatment on Compliers (ToC)	0.072		0.008		-0.069 ***		-0.056 **	
Treatment on Unique Complier (TUC)	0.270 ***		0.009		-0.145 ***		-0.126 **	
Spillover on Compliers (SC)	-0.197 **		-0.001		0.076		0.070	
Spillover on Non-Treated (SNT)	-0.088		-0.015		-0.004		-0.018	
Total Causal Effect (TCE)	-0.002		-0.003		-0.039 **		-0.036 *	

Test scores are standardized to mean zero and standard deviation one in the control group. Odd-numbered columns are OLS regressions, even-numbered columns are IV, using Round 3 data with robust standard errors clustered at the EA level. All regressions except for the TCE are weighted with both sampling and saturation weights to make the results representative of the target population in the study EAs; TCE regressions use sampling weights only. TCE estimated through separate regression of outcomes on a dummy for treatment at the EA level. Baseline values of the following variables are included as controls: age dummies, strata dummies, household asset index, highest grade attended, and an indicator for ever had sex. Significance levels for cross-equation F-tests calculated using multiple equation two-step GMM estimation. Parameter estimates statistically different than zero at 99 percent (***), 95 percent (**), and 90 percent (*) confidence.

TABLE 2. Granular Spillover Analysis

	Dependent Variable:				
	True Saturation in Cell	Terms Enrolled	Average Test Score	Ever Married	Ever Pregnant
	(1)	(2)	(3)	(4)	(5)
CCT 33%	0.244	0.179	0.050	0.021	0.086
	(0.0167)***	(0.0846)**	(0.0119)***	(0.056)	(0.0493)*
CCT 66%	0.372	0.185	0.019	-0.041	0.007
	(0.0487)***	(0.0548)***	(0.00881)**	(0.0244)*	(0.043)
CCT 100%	0.657	0.084	0.019	0.004	0.025
	(0.101)***	(0.063)	(0.012)	(0.021)	(0.040)
UCT 33%	0.246	0.195	0.007	-0.105	-0.082
	(0.0208)***	(0.0585)***	(0.029)	(0.0305)***	(0.0380)**
UCT 66%	0.507	-0.003	0.012	-0.076	-0.089
	(0.0452)***	(0.085)	(0.017)	(0.0348)**	(0.0386)**
UCT 100%	0.691	0.022	0.007	-0.038	-0.022
	(0.0674)***	(0.070)	(0.019)	(0.035)	(0.026)
Spillover CCT 0%	-0.011	0.000	0.001	0.011	0.010
	(0.030)	(0.085)	(0.014)	(0.035)	(0.032)
Spillover CCT 33%	0.247	0.027	0.010	-0.004	-0.038
	(0.0163)***	(0.056)	(0.008)	(0.038)	(0.033)
Spillover CCT 66%	0.381	0.036	0.066	0.017	0.045
	(0.0490)***	(0.073)	(0.0376)*	(0.029)	(0.053)
Spillover UCT 33%	0.244	-0.058	-0.032	0.009	0.013
	(0.0207)***	(0.063)	(0.029)	(0.038)	(0.035)
Spillover UCT 66%	0.510	-0.138	0.015	-0.027	-0.071
	(0.0447)***	(0.163)	(0.039)	(0.046)	(0.0389)*
Observations	2,650	2,579	2,612	2,649	2,650
R-Squared	0.820	0.099	0.418	0.144	0.200

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 with both sampling and saturation weights 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 percent (***), 95 percent (**), and 90 percent (*) confidence.

A Mathematical Appendix

A.1 Preliminary Calculations

This section provides background material used in Theorems 1, 2, 3 and 5.

A.1.1 Form of the MDE

The MDE depends on the standard error of $\widehat{\beta}$:

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

To compute the MDE, we need to determine $SE\left(\widehat{\beta}\right)$. This depends on the data generating process and the randomization structure. Consider a model with a random effects error structure:

$$y_{ic} = \mathbf{x}'_{ic}\beta + v_c + w_{ic}$$

for a vector of treatment status covariates $\mathbf{x}_{ic}$, where v_c is the common cluster component of error and w_{ic} is the individual error. Let $X'_cX_c = \sum_{i=1}^{n} x_{ic}x'_{ic}$ and $u'_c = \begin{bmatrix} u_{1c} & ... & u_{nc} \end{bmatrix}$, where $u_{ic} = v_c + w_{ic}$. Then the standard error of $\widehat{\beta}$ is:

$$SE\left(\widehat{\beta}\right) = \sqrt{\frac{1}{nC} * A^{-1}BA^{-1}}$$

where

$$A := prob\lim \frac{1}{nC}\sum_{c=1}^{C} X'_cX_c \qquad \text{and} \qquad B := prob\lim \frac{1}{nC}\sum_{c=1}^{C} X'_cu_cu'_cX_c$$

Given that all clusters are identical ex-ante, $\frac{1}{N}\sum_{c=1}^{C} E\left[X'_cX_c\right] = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{n}E\left[X'_cX_c\right] = \frac{1}{n}E\left[X'_cX_c\right]$. Also note that A and B are independent of whether one takes $n \to \infty$ or $C \to \infty$. Therefore, using the formulas for matrices A and B yields:

$$A = \frac{1}{n}E[X'_cX_c] \qquad \text{and} \qquad B = \frac{1}{n}E\left[X'_cu_cu'_cX_c\right]$$

The matrix B can be decomposed into two matrices, $B = (n-1)\tau^2 D + (\tau^2 + \sigma^2)A$, which leads to the expression:

$$SE\left(\widehat{\beta}\right) = \sqrt{\left(\frac{n-1}{nC}\right)\tau^2 A^{-1}DA^{-1} + \left(\frac{1}{nC}\right)(\tau^2 + \sigma^2)A^{-1}}$$

where A and D depend on the RS design. We will utilize this expression to calculate $SE\left(\widehat{\beta}\right)$ for different effects and RS designs.

A.1.2 Form of Matrices

$$u_c u_c' = \begin{bmatrix} u_{1c}^2 & u_{1c}u_{2c} & & u_{1c}u_{nc} \\ u_{1c}u_{2c} & \dots & & \dots \\ \dots & & \dots & \\ u_{1c}u_{nc} & & & u_{nc}^2 \end{bmatrix}$$

$$E[u_c u_c'] = \begin{bmatrix} \tau^2+\sigma^2 & \tau^2 & & \tau^2 \\ \tau^2 & \tau^2+\sigma^2 & & \dots \\ \dots & & \dots & \\ \tau^2 & & & \tau^2+\sigma^2 \end{bmatrix} = \tau^2 + \begin{bmatrix} \sigma^2 & 0 & & 0 \\ 0 & \sigma^2 & & \dots \\ \dots & & \dots & \\ 0 & & & \sigma^2 \end{bmatrix}$$

Given $x_{ic}' = \begin{bmatrix} 1 & T_{ic} & S_{ic} & \dots \end{bmatrix}$, (a $1 \times k$ vector), we can write:

$$X_c = \begin{bmatrix} x_{1c}' \\ x_{2c}' \\ \dots \\ x_{nc}' \end{bmatrix} = \begin{bmatrix} 1 & T_{1c} & S_{1c} & \dots \\ 1 & T_{2c} & S_{2c} & \dots \\ \dots & \dots & & \\ 1 & T_{nc} & S_{nc} & \dots \end{bmatrix} \text{ an } n \times k \text{ matrix}$$

$$X_c' X_c = \sum_{i=1}^{n} x_{ic} x_{ic}' = \sum_{i=1}^{n} \begin{bmatrix} 1 & T_{ic} & S_{ic} & \dots \\ T_{ic} & T_{ic}^2 & T_{ic}S_{ic} & \dots \\ S_{ic} & T_{ic}S_{ic} & S_{ic}^2 & \dots \\ \dots & & & \dots \end{bmatrix} \text{ a } k \times k \text{ matrix}$$

A.1.3 Relevant Expectations

Distribution of treatment status: [46]

- $E[T_{ic}] = P(T_{ic} = 1) = \mu$
- $E[S_{ic}] = 1 - \mu - \psi = \mu_S$
- $E[C_{ic}] = \psi$
- $E[T_{ic}^x] = E[T_{ic}] = \mu$

[46] We implicitly assume that realized saturation is equal to assigned saturation, i.e. $\pi_c = \frac{1}{n}\Sigma_{i=1}^{n} T_{ic}$, so given assigned saturation, there is no variation in realized saturation.

- $E[S_{ic}^x] = E[S_{ic}] = \mu_S$
- $E\left[T_{ic}S_{ic}\right] = 0$

Variance of treatment status:

- $Var[T_{ic}] = E[T_{ic}^2] - E[T_{ic}]^2 = \mu(1-\mu)$
- $Var[S_{ic}] = (1-\mu-\psi)(\mu+\psi)$
- $Var[C_{ic}] = \psi(1-\psi)$

Within cluster treatment status:

- $E\left[T_{ic}T_{jc}\right] = P(T_{ic}=1, T_{jc}=1) = \sum_{\Pi} P(T_{ic}=1, T_{jc}=1, \pi_c=\pi) = \sum_{\Pi} \pi^2 f(\pi) = E[\pi^2]$

  where the second equality follows from the chain rule of probability and the third 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 π_c.

- $E\left[S_{ic}S_{jc}\right] = 1 - 2\mu + E[\pi^2] - \psi$

$$\begin{aligned} E\left[S_{ic}S_{jc}\right] &= P(S_{ic}=1, S_{jc}=1) \\ &= \sum_{\Pi} P(S_{ic}=1, S_{jc}=1, T_c=\pi) \\ &= \sum_{\Pi\setminus\{0\}} (1-\pi)^2 f(\pi) \\ &= E\left[(1-\pi)^2\right] - \psi \\ &= 1 - 2\mu + E[\pi^2] - \psi \end{aligned}$$

- $E\left[C_{ic}C_{jc}\right] = \psi$
- $E\left[T_{ic}S_{jc}\right] = \mu - E[\pi^2]$

$$\begin{aligned} E\left[T_{ic}S_{jc}\right] &= P(T_{ic}=1, S_{jc}=1) \\ &= \sum_{\Pi} P(T_{ic}=1, S_{jc}=1, T_c=\pi) \\ &= \sum_{\Pi\setminus\{0\}} \pi(1-\pi) f(\pi) \\ &= E\left[\pi(1-\pi)\right] - 0 * \psi \\ &= \mu - E[\pi^2] \end{aligned}$$

Across cluster treatment status:

- $E\left[T_{ic}T_{jd}\right] = \mu^2$
  since $E\left[T_{ic}T_{jd}\right] = E\left[T_{ic}\right]E\left[T_{id}\right]$ by independence
- $E\left[S_{ic}S_{jd}\right] = \left(1-\mu-\psi\right)^2$
- $E\left[C_{ic}C_{jd}\right] = \psi^2$

Correlation with saturation π_c is

- $E[T_{ic}\pi_c^x] = \sum_{\Pi}\pi^x P(T_{ic}=1,\pi_c=\pi) = \sum_{\Pi}\pi^{x+1}f\left(\pi\right) = E[\pi^{x+1}]$
- $E[T_{ic}^x\pi_c] = E[T_{ic}\pi_c]$
- $E[T_{ic}T_{jc}\pi_c^x] = \sum_{\Pi}\pi^x P(T_{ic}=1,T_{jc}=1,\pi_c=\pi) = \sum_{\Pi}\pi^{x+2}f\left(\pi\right) = E[\pi^{x+2}]$
- $E[S_{ic}\pi_c^x] = \sum_{\Pi}\pi^x P(S_{ic}=1,\pi_c=\pi) = \sum_{\Pi\setminus\{0\}}\pi^x\left(1-\pi\right)f\left(\pi\right) = E\left[(1-\pi)\pi^x\right]$
- $E[S_{ic}^x\pi_c] = E[S_{ic}\pi_c]$
- $E\left[S_{ic}S_{jc}\pi_c^x\right] = \sum_{\Pi}\pi^x P(S_{ic}=1,S_{jc}=1,\pi_c=\pi) = \sum_{\Pi}\pi^x\left(1-\pi\right)^2 f\left(\pi\right) = E[\pi^x\left(1-\pi\right)^2]$
- $E[T_{ic}S_{jc}\pi_c^x] = \sum_{\Pi}\pi^x P(T_{ic}=1,S_{jc}=1,\pi_c=\pi) = \sum_{\Pi}\pi^x\pi\left(1-\pi\right)f\left(\pi\right) = E[\pi^{x+1}\left(1-\pi\right)]$

Correlation of treatment status between two girls in the same cluster:

$$\begin{aligned}
\rho_T &= \frac{E\left[T_{ic}T_{jc}\right]-E\left[T_{ic}\right]E\left[T_{ic}\right]}{Var[T_{ic}]} = \frac{\eta^2}{\mu(1-\mu)} \\
\rho_S &= \frac{E\left[S_{ic}S_{jc}\right]-E\left[S_{ic}\right]E\left[S_{ic}\right]}{Var[S_{ic}]} \\
&= \frac{\left(1-2\mu+\eta^2+\mu^2-\psi\right)-\left(1-\mu-\psi\right)^2}{\left(1-\mu-\psi\right)\left(\mu+\psi\right)} \\
&= \frac{\eta^2+\psi\left(1-2\mu-\psi\right)}{\left(1-\mu-\psi\right)\left(\mu+\psi\right)} \\
\rho_C &= 1
\end{aligned}$$

Distribution of u_{ic}:

- $E\left[u_{ic}^2\right] = \tau^2+\sigma^2$
- $E\left[u_{ic}u_{jc}\right] = \tau^2$ if $i \neq j$ which is $Cov(u_{ic}u_{jc})$
- $E\left[u_{ic}u_{jd}\right] = 0$ if $c \neq d$

- T_{ic} or S_{ic} is independent of $u_{ic} \Rightarrow E\left[f(u_{ic})g(T_{ic})\right] = E\left[f(u_{ic})\right] * E\left[g(T_{ic})\right]$

Sum of the error terms and treatment status within each cluster:

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

since T_{ic} and S_{ic} are independent of u_{ic}.

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

since $T_{ic}S_{ic} = 0$ for all i, c.

A.1.4 Variance of Treatament Saturation

The marginal distribution of saturations across treatment clusters (removing control clusters) is:

$$g(\pi) = \frac{f(\pi)}{1 - \psi}$$

with support $\Pi\backslash\{0\}$.

$$\begin{aligned}
E_g[\pi^2] &= \frac{1}{1-\psi}\sum_{\Pi\backslash\{0\}} \pi^2 f(\pi) \\
E_g[\pi] &= \frac{1}{1-\psi}\sum_{\Pi\backslash\{0\}} \pi f(\pi)
\end{aligned}$$

The component of total variation in cluster saturation due to variation in the saturation of treated clusters is:

$$\begin{aligned}\eta_T^2: &= Var(\pi|\pi>0) = E_g[\pi^2] - E_g[\pi]^2 \\ &= \left(\frac{1}{1-\psi}\right)\left[\sum_{\Pi\setminus\{0\}} \pi^2 f(\pi) - \left(\frac{1}{1-\psi}\right)\left(\sum_{\Pi\setminus\{0\}} \pi f(\pi)\right)^2\right] \\ &= \left(\frac{1}{1-\psi}\right)\eta^2 - \left(\frac{\psi}{(1-\psi)^2}\right)\mu^2\end{aligned}$$

Then the total variation can be expressed as the sum of the variation in the saturation of treated clusters and the variation between treated and control clusters, weighted by the size of the control group:

$$\eta^2 = (1-\psi)\,\eta_T^2 + \left(\frac{\psi}{1-\psi}\right)\mu^2$$

A.2 Proof of Theorem 1

We want to compute matrices A and B for the model with $x_{ic}' = \begin{bmatrix} 1 & T_{ic} & S_{ic} \end{bmatrix}$. Using the calculations from Section A.1, we can calculate:

$$A = \frac{1}{n}\sum_{i=1}^{n} E\begin{bmatrix} n & T_{ic} & S_{ic} \\ T_{ic} & T_{ic}^2 & T_{ic}S_{ic} \\ S_{ic} & T_{ic}S_{ic} & S_{ic}^2 \end{bmatrix} = \begin{bmatrix} 1 & \mu & \mu_S \\ \mu & \mu & 0 \\ \mu_S & 0 & \mu_S \end{bmatrix}$$

$$\begin{aligned} B &= \frac{1}{n}E\begin{bmatrix} \left(\sum_{i=1}^n u_{ic}\right)^2 & \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n T_{ic}u_{ic}\right) & \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n S_{ic}u_{ic}\right) \\ \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n T_{ic}u_{ic}\right) & \left(\sum_{i=1}^n T_{ic}u_{ic}\right)^2 & \left(\sum_{i=1}^n T_{ic}u_{ic}\right)\left(\sum_{i=1}^n S_{ic}u_{ic}\right) \\ \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n S_{ic}u_{ic}\right) & \left(\sum_{i=1}^n T_{ic}u_{ic}\right)\left(\sum_{i=1}^n S_{ic}u_{ic}\right) & \left(\sum_{i=1}^n S_{ic}u_{ic}\right)^2 \end{bmatrix} \\ &= (n-1)\tau^2\begin{bmatrix} 1 & \mu & \mu_S \\ \mu & \eta^2+\mu^2 & \mu-\mu^2-\eta^2 \\ \mu_S & \mu-\mu^2-\eta^2 & \mu_S-\mu+\eta^2+\mu^2 \end{bmatrix} + \left(\tau^2+\sigma^2\right)A \end{aligned}$$

Using mathematica to compute $SE\left(\widehat{\beta}\right) = \sqrt{\frac{1}{nC} * A^{-1}BA^{-1}}$, taking the diagonal entries and plugging in the expression relating η^2 and η_T^2 yields the result. Q.E.D.

Proof of Corollary 1: Fixing μ and ψ, $SE\left(\widehat{\beta}_1\right)$ and $SE\left(\widehat{\beta}_2\right)$ are both minimized at $\eta_T^2 = 0$. This corresponds to a partial population experiment with a control group of size ψ and a treatment saturation of $P = \mu/(1-\psi)$. Q.E.D.

Proof of Corollary 2: Fixing ψ, a partial population design has the smallest sum of standard errors, for any treatment size μ. Therefore, we can restrict attention to the set of partial population designs, and the expression for $Var\left(\widehat{\beta}\right)$ simplifies to:

$$MDE_{\omega}^{T} = [t_{1-\gamma} + t_{\alpha}]\sqrt{\frac{1}{nC} * \left\{(n-1)\,\tau^2\left(\frac{1}{(1-\psi)\,\psi}\right) + (\tau^2+\sigma^2)\left(\frac{\psi+\mu}{\mu\psi}\right)\right\}}$$

$$MDE_{\omega}^{S} = [t_{1-\gamma} + t_{\alpha}]\sqrt{\frac{1}{nC} * \left\{(n-1)\,\tau^2\left(\frac{1}{(1-\psi)\,\psi}\right) + (\tau^2+\sigma^2)\left(\frac{1-\mu}{(1-\mu-\psi)\psi}\right)\right\}}$$

The sum of these expressions is minimized at $\mu = \mu_S = (1-\psi)/2$, which corresponds to a partial population experiment with $P = 1/2$. Q.E.D.

Proof of Corollary 3: In a partial population design with $\mu = (1-\psi)/2$,

$$MDE_{\omega}^{T} = MDE_{\omega}^{S} = [t_{1-\gamma} + t_{\alpha}]\sqrt{\frac{1}{nC} * \left\{(n-1)\,\tau^2\left(\frac{1}{(1-\psi)\,\psi}\right) + (\tau^2+\sigma^2)\left(\frac{\psi+1}{(1-\psi)\psi}\right)\right\}}$$

When there is no inter-cluster correlation, $\tau^2 = 0$, the expression simplifies to:

$$[t_{1-\gamma} + t_{\alpha}]\sqrt{\frac{1}{nC} * \sigma^2\left(\frac{\psi+1}{(1-\psi)\psi}\right)}$$

which is minimized at $\psi^* = \sqrt{2}-1$. When there is no individual error, $\sigma^2 = 0$, the expression simplifies to:

$$[t_{1-\gamma} + t_{\alpha}]\sqrt{\frac{1}{nC} * \left\{\tau^2\left(\frac{n+\psi}{(1-\psi)\,\psi}\right)\right\}}$$

which is minimized at $\psi^* = \sqrt{n(1+n)} - n$. Note $\lim_{n\to\infty}\sqrt{n(1+n)} - n = 1/2$. Given that $(\psi+1)/((1-\psi)\psi)$ and $(\psi+n)/((1-\psi)\psi)$ are both convex with unique minimums, any weighted sum of these functions is minimized at a value ψ^* that lies between the minimum of each function. Therefore, when $\tau^2 > 0$ and $\sigma^2 > 0$, $\psi^* \in (\sqrt{2}-1, \sqrt{n(1+n)} - n)$.Q.E.D.

Proof of Corollary 4: Follows directly from Theorem 1. Q.E.D.

A.3 Proof of Theorem 2

We want to compute matrices A and B for the model with $x'_{ic} = \begin{bmatrix} 1 & T_{1ic} & S_{1ic} & T_{2ic} & S_{2ic} \end{bmatrix}$ where $T_{1ic} = \mathbb{1}(T_{ic} = 1, \pi_c = \pi_1)$, $S_{1ic} = \mathbb{1}(T_{ic} = 0, \pi_c = \pi_1)$, $T_{2ic} = \mathbb{1}(T_{ic} = 1, \pi_c = \pi_{2ic})$ and so forth. Using the calculations from Section A.1 and defining $\mu_k := \pi_k f(\pi_k)$, $p_k := (1 - \pi_k) f(\pi_k)$, $\eta_k := \pi_k^2 f(\pi_k)$ and $\rho_k := (1 - \pi_k)^2 f(\pi_k) = p_k - \mu_k + \eta_k$, we can calculate:

$$A = \frac{1}{n} \sum_{i=1}^{n} E \begin{bmatrix} n & T_{1ic} & S_{1ic} & T_{2ic} & S_{2ic} \\ T_{1ic} & T_{1ic}^2 & 0 & 0 & 0 \\ S_{1ic} & 0 & S_{1ic}^2 & 0 & 0 \\ T_{2ic} & 0 & 0 & T_{2ic}^2 & 0 \\ S_{2ic} & 0 & 0 & 0 & S_{2ic}^2 \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & p_1 & \mu_2 & p_2 \\ \mu_1 & \mu_1 & 0 & 0 & 0 \\ p_1 & 0 & p_1 & 0 & 0 \\ \mu_2 & 0 & 0 & \mu_2 & 0 \\ p_2 & 0 & 0 & 0 & p_2 \end{bmatrix}$$

$$\begin{aligned} B &= \frac{1}{n} E \begin{bmatrix} (\sum_{i=1}^n u_{ic}) \\ (\sum_{i=1}^n T_{1ic} u_{ic}) \\ (\sum_{i=1}^n S_{1ic} u_{ic}) \\ (\sum_{i=1}^n T_{2ic} u_{ic}) \\ (\sum_{i=1}^n S_{2ic} u_{ic}) \end{bmatrix} * \begin{bmatrix} (\sum_{i=1}^n u_{ic}) \\ (\sum_{i=1}^n T_{1ic} u_{ic}) \\ (\sum_{i=1}^n S_{1ic} u_{ic}) \\ (\sum_{i=1}^n T_{2ic} u_{ic}) \\ (\sum_{i=1}^n S_{2ic} u_{ic}) \end{bmatrix}' \\ &= (n-1)\tau^2 \begin{bmatrix} 1 & \mu_1 & p_1 & \mu_2 & p_2 \\ \mu_1 & \eta_1 & \mu_1 - \eta_1 & 0 & 0 \\ p_1 & \mu_1 - \eta_1 & \rho_1 & 0 & 0 \\ \mu_2 & 0 & 0 & \eta_2 & \mu_2 - \eta_2 \\ p_2 & 0 & 0 & \mu_2 - \eta_2 & \rho_2 \end{bmatrix} + \left(\tau^2 + \sigma^2\right) A \end{aligned}$$

Using mathematica to compute $SE\left(\widehat{\beta}\right) = \sqrt{\frac{1}{nC} * A^{-1} B A^{-1}}$ and taking the diagonal entries yields the MDE^T for each saturation π_j:

$$\begin{aligned} MDE_\omega^T(\pi_j) &= (t_{1-\gamma} + t_\alpha) \sqrt{\frac{1}{nC} * \left\{ (n-1)\tau^2 \left(\frac{\eta_j}{\mu_j^2} + \frac{1}{\psi} \right) + (\tau^2 + \sigma^2) \left(\frac{\psi + \mu_j}{\psi \mu_j} \right) \right\}} \\ &= (t_{1-\gamma} + t_\alpha) \sqrt{\frac{1}{nC} * \left\{ (n-1)\tau^2 \left(\frac{1}{f(\pi_j)} + \frac{1}{\psi} \right) + (\tau^2 + \sigma^2) \left(\frac{1}{\mu_j} + \frac{1}{\psi} \right) \right\}} \end{aligned}$$

Next, we can compute $MDSE_\omega^T(\pi_j, \pi_k)$ from

$$SE(\delta_{jk}^T) = SE[\left(\beta_{1\pi_k} - \beta_{1\pi_j}\right) / (\pi_k - \pi_j)] = SE\left(\beta_{1\pi_k} - \beta_{1\pi_j}\right) / (\pi_k - \pi_j)$$

as

$$\begin{aligned} Cov(\beta_{1\pi_k}, \beta_{1\pi_j}) &= \frac{1}{nC}(n\tau^2 + \sigma^2)\frac{1}{\psi} \\ Var(\beta_{1\pi_k} - \beta_{1\pi_j}) &= Var(\beta_{1\pi_j}) + Var(\beta_{1\pi_k}) - 2Cov(\beta_{1\pi_k}, \beta_{1\pi_j}) \\ &= \frac{1}{nC} * \left\{ (n-1)\,\tau^2 \left(\frac{1}{f(\pi_j)} + \frac{1}{f(\pi_k)} \right) + \left(\tau^2 + \sigma^2\right) \left(\frac{1}{\mu_j} + \frac{1}{\mu_k} \right) \right\} \end{aligned}$$

Plugging this into the expression for $MDSE^T$ yields the result. Calculating the MDE^S and $MDSE^S$ is analogous. Q.E.D.

Proof of Corollary 5: Fixing the size of each saturation bin $f(\pi_j) = f_j$ and $f(\pi_k) = f_k$ and the distance between two saturations $\pi_k - \pi_j = \Delta$, minimizing $MDSE^T_\omega(\pi_j, \pi_k)$ $+MDSE^S_\omega(\pi_j, \pi_k)$ is equivalent to solving:

$$\min_{\pi_j} \left(\frac{1}{f_j \pi_j} + \frac{1}{f_k(\pi_j + \Delta)} + \frac{1}{f_j(1 - \pi_j)} + \frac{1}{f_k(1 - \Delta - \pi_j)} \right)$$

The minimum occurs at the π_j^* that solves $\pi_j^*(1 - \pi_j^*)f_j = (\pi_j^* + \Delta)(1 - \Delta - \pi_j^*)f_k$. When $f_j = f_k$, $\pi_j^* = (1 - \Delta)/2$ and $\pi_k^* = \pi_j^* + \Delta = (1 + \Delta)/2$, which is symmetric about 1/2.

Fixing $f_j = f_k$, the Δ that minimizes $MDSE^T_\omega(\pi_j, \pi_k) + MDSE^S_\omega(\pi_j, \pi_k)$ is equivalent to solving:

$$\min_{\Delta} \frac{1}{\Delta^2} \left(\frac{(n-1)}{n}\tau^2 + \frac{(\tau^2 + \sigma^2)}{n} \left(\frac{2}{(1 - \Delta)(1 + \Delta)} \right) \right)$$

The optimal Δ^* solves:

$$\frac{(n-1)\tau^2}{2(\tau^2 + \sigma^2)} = \frac{2(\Delta^*)^2 - 1}{(1 - (\Delta^*)^2)^2}$$

If $\tau^2 = 0$, then $2(\Delta^*)^2 - 1 = 0$, yielding $\Delta^* = \sqrt{2}/2$. Note that $(2\Delta^2 - 1)/((1 - \Delta^2)^2)$ is monotonically increasing for $\Delta \in [0, 1)$, and strictly positive for $\Delta > \sqrt{2}/2$. When $\tau > 0$, $((n-1)\tau^2)(2(\tau^2 + \sigma^2))$ is also strictly positive, increasing in τ^2 and decreasing in σ^2. Therefore, $\Delta^* \in \left(\sqrt{2}/2, 1\right)$ for $\tau^2 > 0$ and finite n, Δ^* is increasing in τ^2 and n, and decreasing in σ^2. If $\tau^2 > 0$, then the left hand side converges to ∞ as $n \to \infty$, which requires $\Delta^* \to 1$. Q.E.D.

A.4 Proof of Theorem 3

We want to compute matrices A and B for the model with

$$x'_{ic} = \begin{bmatrix} 1 & T_{ic} & T_{ic}\pi_c & S_{ic} & S_{ic}\pi_c \end{bmatrix}$$

Using the calculations from Section A.1, we can calculate:

$$\begin{aligned} A &= \frac{1}{n}\sum_{i=1}^{n} E \begin{bmatrix} 1 & T_{ic} & T_{ic}\pi_c & S_{ic} & S_{ic}\pi_c \\ T_{ic} & T_{ic}^2 & T_{ic}^2\pi_c & T_{ic}S_{ic} & T_{ic}S_{ic}\pi_c \\ T_{ic}\pi_c & T_{ic}^2\pi_c & T_{ic}^2\pi_c^2 & T_{ic}S_{ic}\pi_c & T_{ic}S_{ic}\pi_c^2 \\ S_{ic} & T_{ic}S_{ic} & T_{ic}S_{ic}\pi_c & S_{ic}^2 & S_{ic}^2\pi_c \\ S_{ic}\pi_c & T_{ic}S_{ic}\pi_c & T_{ic}S_{ic}\pi_c^2 & S_{ic}^2\pi_c & S_{ic}^2\pi_c^2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \mu & \eta^2+\mu^2 & 1-\mu-\psi & \mu-\eta^2+\mu^2 \\ \mu & \mu & \eta^2+\mu^2 & 0 & 0 \\ \eta^2+\mu^2 & \eta^2+\mu^2 & E[\pi^3] & 0 & 0 \\ 1-\mu-\psi & 0 & 0 & 1-\mu-\psi & \mu-\eta^2+\mu^2 \\ \mu-\eta^2+\mu^2 & 0 & 0 & \mu-\eta^2+\mu^2 & \eta^2+\mu^2-E[\pi^3] \end{bmatrix} \end{aligned}$$

$$\begin{aligned} B &= \frac{1}{n} E \begin{bmatrix} (\sum_{i=1}^n u_{ic}) \\ (\sum_{i=1}^n T_{ic}u_{ic}) \\ (\sum_{i=1}^n T_{ic}\pi_c u_{ic}) \\ (\sum_{i=1}^n S_{ic}u_{ic}) \\ (\sum_{i=1}^n S_{ic}\pi_c u_{ic}) \end{bmatrix} * \begin{bmatrix} (\sum_{i=1}^n u_{ic}) \\ (\sum_{i=1}^n T_{ic}u_{ic}) \\ (\sum_{i=1}^n T_{ic}\pi_c u_{ic}) \\ (\sum_{i=1}^n S_{ic}u_{ic}) \\ (\sum_{i=1}^n S_{ic}\pi_c u_{ic}) \end{bmatrix}^{\intercal} \\ &= (n-1)\tau^2 D + \left(\tau^2+\sigma^2\right) A \end{aligned}$$

where

$$D = \begin{bmatrix} 1 & \mu & E[\pi^2] & 1-\mu-\psi & \mu-E[\pi^2] \\ \mu & E[\pi^2] & E[\pi^3] & \mu-E[\pi^2] & E[\pi^2]-E[\pi^3] \\ E[\pi^2] & E[\pi^3] & E[\pi^4] & E[\pi^2]-E[\pi^3] & E[\pi^3]-E[\pi^4] \\ 1-\mu-\psi & \mu-E[\pi^2] & E[\pi^2]-E[\pi^3] & 1-2\mu+E[\pi^2]-\psi & \mu-2E[\pi^2]+E[\pi^3] \\ \mu-E[\pi^2] & E[\pi^2]-E[\pi^3] & E[\pi^3]-E[\pi^4] & \mu-2E[\pi^2]+E[\pi^3] & E[\pi^2]-2E[\pi^3]+E[\pi^4] \end{bmatrix}$$

Using mathematica to compute $SE\left(\widehat{\delta}\right) = \sqrt{\frac{1}{nC} * A^{-1}BA^{-1}}$ and taking the diagonal entries yields the result. The $MDSE^T$ is a function of $SE(\widehat{\delta}_3)$, while the $MDSE^S$ is a function of

$SE(\widehat{\delta_4})$.

B Additional Analysis

This section presents additional possibilities that an RS design can identify. First, we illustrate how an RS design can consistently estimate the pure control outcome, a useful result for situations in which institutional constraints prohibit including pure control outcomes. Second, we compute the power of an RS design to detect treatment effects when it is determined ex post that there are no spillover effects. Third, we show that an RS design can also be used to estimate spillover effects in overlapping networks, such as a friends or family network.

B.1 Inference Without a Pure Control

The RS design opens up unique empirical possibilities even when there is no pure control group. This is particularly important for settings in which a pure control is not feasible due to regulatory requirements or other exogenous restrictions.[47] If a study has no pure control group, the counterfactual is at the mercy of within-cluster spillovers. In this context, the RS design has the distinct advantage of allowing a researcher to test for the presence of spillover effects and estimate the unperturbed counterfactual. If the spillover effect is continuous at zero, the researcher can use the variation in treatment saturation to project what would happen to untreated individuals as the saturation approaches zero.[48] With this unperturbed counterfactual in hand, we can then estimate $\overline{\hat{SNT}}$, and use this value to correct the estimate of the $\overline{\hat{ITT}}$.

Assumption 5 provides a simple way to estimate the pure control by assuming that the outcome variable is linear with respect to treatment saturation.

Assumption 5 (Linearity)**.** $E(Y|T,\pi)$ *is an affine function of* π*.*

While it is possible to use a more flexible functional form and the specification can be tested, the linear case provides simple intuition for the technique.[49] Given Assumption 5, it

[47] For example, in McIntosh et al. (2013), a Mexican government rule required that each participating cluster (municipality) be guaranteed at least one treated sub-unit (neighborhood).

[48] Although continuity is a reasonable assumption, it is not universally applicable. Consider signalling in a ground-hog colony. Individuals are 'treated' by being alerted to the presence of a nearby predator, and the possible individual-level outcomes are 'aware' and 'not aware'. The animal immediately signals danger to the rest of the colony, and control outcomes will be universally 'aware' for any positive treatment saturation, but 'unaware' when the saturation is exactly zero.

[49] In a panel difference in difference regression, the quantity giving the desired counterfactual would be the un-interacted 'post-treatment' dummy. This is the change the control group would have experienced at saturation zero

is natural to estimate:

$$Y_{ic} = \delta_0 + \delta_1 T_{ic} + \delta_2 * \pi_c + \delta_3 (T_{ic} * \pi_c) + \varepsilon_{ic} \quad (5)$$

Given RS design ω with no pure control, estimating Equation 6 with saturation weights and Equation 5, the hypothesis test $\delta_2 = 0$ determines whether there is variation in the control outcome across saturations. If spillovers are present on untreated individuals, then the counterfactual needs to be corrected. The coefficient $\hat{\delta}_0$ is an estimate of the desired 'pure' control outcome, $E(Y_{ic} \mid T_{ic} = 0, \pi_c = 0)$, while $\hat{\beta}_0$ is an estimate of the within-cluster control outcome actually used as the counterfactual, $E(Y_{ic} \mid T_{ic} = 0, \pi_c > 0)$. The difference between $\hat{\beta}_0$ and $\hat{\delta}_0$ is the $\overline{\hat{SNT}}$, which can be used to derive an unbiased estimate of the $\overline{\hat{ITT}}$.

Theorem 4 (Consistency with No Control)**.** *Assume Assumption 1, 2, 3 and 5, and let ω be a randomized saturation design with no pure control and $\kappa \geq 2$ interior saturations. Then ω generates consistent estimates $\overline{\hat{ITT}}_\omega = \hat{\beta}_1 + \hat{\beta}_0 - \hat{\delta}_0$ and $\overline{\hat{SNT}}_\omega = \hat{\beta}_0 - \hat{\delta}_0$, where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\delta}_0$ are OLS estimates from equation 6 with saturation weights and equation 5.*

Proof. Given Assumption 5, we can identify the slope of the ITT and SNT. The rest of the proof follows easily from the Law of Large Numbers. □

Similar estimates for the ITT and SNT at a specific saturation are generated by estimating Equation 6 on a single saturation.

B.2 Using Within-cluster Controls as Counterfactuals

Suppose there is no evidence of spillovers on untreated individuals – the estimate of $SNT(\pi)$ is a precise zero for all π.[50]

Assumption 6. $SNT(\pi) = 0$ *for all* $\pi \in \Pi$.

Then the within-cluster controls are not subject to interference from the treatment and they can be used as counterfactuals to increase the power of the treatment effect estimates. The researcher can pool within-cluster and pure controls, and estimate a simpler model to measure treatment effects:

$$Y_{ic} = \beta_0 + \beta_1 T_{ic} + \varepsilon_{ic} \quad (6)$$

[50]This assumption is testable using any RS design that yields a consistent estimate of $\hat{SNT}(\pi)$.

Given RS design ω, this regression returns $\widehat{\overline{ITT}}_\omega = \hat{\beta}_1$.[51] Power is significantly improved by the larger counterfactual, particularly when τ is high.

Theorem 5 characterizes the pooled MDE when the within-cluster controls are included in the counterfactual.

Theorem 5 (MDE with Within-Cluster Controls)**.** *Assume Assumptions 1, 2, 3 and 6 and let ω be a randomized saturation design. Then the MDE of $\overline{ITT}_\omega$ for statistical significance level α and power γ is:*

$$MDE_\omega^T = (t_{1-\gamma} + t_\alpha)\sqrt{\frac{1}{nC}\left\{\left(\frac{(1+\rho(n-1))}{\mu(1-\mu)}\right)\tau^2 + \left(\frac{1}{\mu(1-\mu)}\right)\sigma^2\right\}}$$

where $\rho = \eta^2/\mu(1-\mu)$ is the correlation in treatment status between two individuals in the same cluster.

Proof. Given the model $x_{ic}' = \begin{bmatrix} 1 & T_{ic} \end{bmatrix}$ and calculations from Appendix A.1, we can calculate:

$$A = \frac{1}{n}E\begin{bmatrix} n & \sum_{i=1}^n T_{ic} \\ \sum_{i=1}^n T_{ic} & \sum_{i=1}^n T_{ic}^2 \end{bmatrix} = \begin{bmatrix} 1 & \mu \\ \mu & \mu \end{bmatrix}$$

and

$$\begin{aligned} B &= \frac{1}{n}E\begin{bmatrix} \left(\sum_{i=1}^n u_{ic}\right)^2 & \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n T_{ic}u_{ic}\right) \\ \left(\sum_{i=1}^n u_{ic}\right)\left(\sum_{i=1}^n T_{ic}u_{ic}\right) & \left(\sum_{i=1}^n T_{ic}u_{ic}\right)^2 \end{bmatrix} \\ &= \tau^2(n-1)\begin{bmatrix} 1 & \mu \\ \mu & \eta^2+\mu^2 \end{bmatrix} + \left(\tau^2+\sigma^2\right)A \end{aligned}$$

This can be used to compute

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

Using $\eta^2 = \rho\mu(1-\mu)$, we can express $SE\left(\widehat{\beta}_1\right)$ in terms of μ and ρ.

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

Fixing μ, this expression is minimized at $\eta^2 = 0$ or $\rho = 0$. Q.E.D.

[51] Saturation weights are necessary if there are spillover effects on treated individuals, $ST(\pi) \neq 0$ for some $\pi \in \Pi$.

□

Theorem 5 nests the MDE of this model between the more familiar expressions for the MDE of the blocked and clustered designs, and provides context for two well-known results. Fixing the treatment probability μ, the expression for the MDE is decreasing in the variance of the treatment saturation η^2, and minimized when this variation is zero, which corresponds to the blocked design. Second, fixing η^2, the MDE is minimized when $\mu(1-\mu)$ is maximized, which occurs at $\mu = 1/2$. Therefore, as is well known, the optimal design in the absence of spillovers is a blocked study with equal size treatment and control groups.

An immediate result of Theorem 5 is that the power of the pooled treatment effect in any RS design lies between the power of the treatment effect in the blocked and clustered designs.

Corollary 6 (Nesting of MDE)**.** *Let ω be a randomized saturation design with treatment probability μ. Then*

$$MDE_B^T < MDE_\omega^T < MDE_C^T,$$

where MDE_B^T is the MDE in a blocked design with saturation μ and MDE_C^T is the MDE in a clustered design with share of treatment clusters μ.

Proof. Follows directly from Theorem 5, noting that the blocked design corresponds to $\rho = 0$ and the clustered design corresponds to $\rho = 1$. Q.E.D. □

B.3 Spillover Effects in Other Networks

The RS design we present must be implemented in a non-overlapping network (such as villages or schools), but many networks of interest do not satisfy this strong requirement (such as peer networks or extended families). However, an RS design implemented on a non-overlapping network also produces exogenous variation in the treatment saturation of overlapping networks, variation that is always superior to what would be obtained from a *blocked* design and generally superior to *clustered* designs. This variation depends on the structure of both networks – it increases as the correlation between the two networks increases. As implementing a RS design using non-overlapping clusters is much more straightforward than the sequential randomization required to conduct a RS design in overlapping networks (Toulis and Kao 2013), this provides an attractive way of generating random variation in treatment saturation even when the true network of interest is overlapping.

Figure 5 illustrates the treatment saturation distributions in an overlapping network that results from implementing either a blocked, clustered or RS design on a non-overlapping

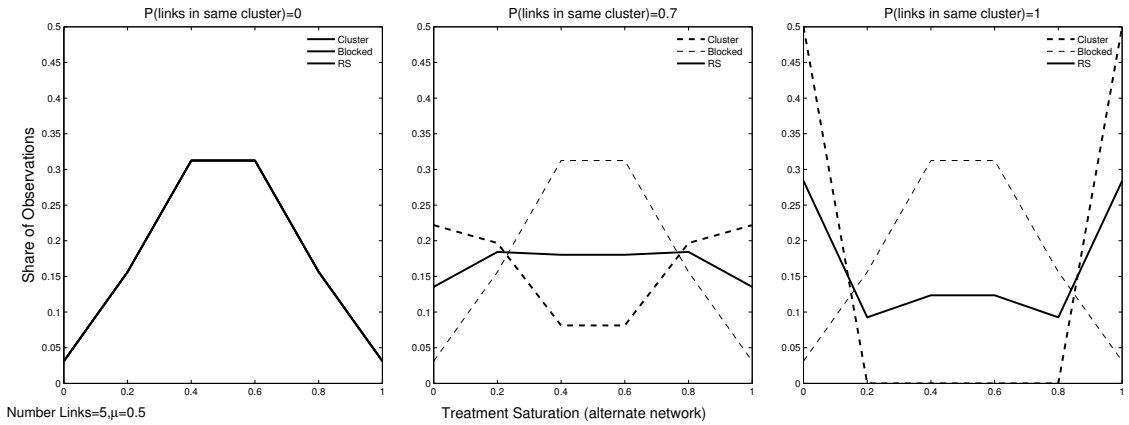

FIGURE 5. Treatment Saturation of Alternate Network

network.[52] Using an overlapping network with five links per individual, we plot the share of individuals at each treatment saturation in the non-overlapping network, where the treatment saturation captures the share of an individual's links who receive treatment. We use the probability that a link in the overlapping network connects two individuals in the same cluster (the unit of the non-overlapping network) to measure the correlation between networks.[53]

As can be seen in Figure 5, the *blocked* design produces little overall variation in treatment saturations; the saturations are centered around 50%, independent of the correlation. The *clustered* design suffers from the opposite problem: because treatment has taken place at the cluster level, it is dominated by nodes that have either high or low treatment saturations when there is correlation between networks. Finally, the RS design produces a more even distribution of saturations when there is correlation between networks. In the limit, when there is no correlation between the two networks, the three designs produce the same saturation distributions (left panel of Figure 5).

[52]We use a blocked design in which 50% of individuals in each cluster are treated, a clustered design in which 50% of clusters are treated at either 100% or 0% saturation, and a RS design in which an equal share of clusters are treated at saturations 0%, 33%, 67% or 100%. Each assignment rule results in the same overall fraction (one half) of the sample being treated.

[53]The specific structure of the network is irrelevant. Any network with the same number of links and correlation measure will acheive the same saturation distribution.

TABLE A1. Balance Tests

PANEL A: Pooled Tests.

	Dependent Variable:									
	Household Size	Asset Index	Female Headed Household	Mobile Phone Ownership	Age	Highest Grade at Baseline	Mother Alive	Father Alive	Never had Sex	Ever Pregnant
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CCT	-0.027	0.330	-0.0889**	-0.034	-0.251**	-0.218	-0.0422*	-0.002	-0.005	0.004
	(0.195)	(0.298)	(0.040)	(0.051)	(0.113)	(0.169)	(0.025)	(0.039)	(0.027)	(0.009)
UCT	0.219	0.464**	-0.0773*	-0.032	0.138	0.355**	-0.001	0.051	-0.017	0.006
	(0.162)	(0.221)	(0.042)	(0.064)	(0.136)	(0.154)	(0.026)	(0.038)	(0.038)	(0.009)
Within CCT EA Control	-0.155	0.078	-0.0923**	0.014	0.115	-0.058	0.011	0.056	0.011	0.007
	(0.152)	(0.326)	(0.036)	(0.048)	(0.130)	(0.192)	(0.027)	(0.034)	(0.030)	(0.011)
Within UCT EA Control	-0.150	0.392*	-0.074	0.055	0.012	0.136	0.015	0.003	0.019	0.005
	(0.236)	(0.234)	(0.049)	(0.059)	(0.170)	(0.171)	(0.036)	(0.057)	(0.025)	(0.021)
Mean in Pure Control:	6.432	0.581	0.343	0.616	15.252	7.479	0.842	0.705	0.797	0.023
Observations	2,651	2,651	2,651	2,651	2,653	2,652	2,653	2,648	2,653	2,652

PANEL B: Linear Slope Term.

	Household Size	Asset Index	Female Headed Household	Mobile Phone Ownership	Age	Highest Grade at Baseline	Mother Alive	Father Alive	Never had Sex	Ever Pregnant
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CCT	-0.691	-1.242**	-0.027	-0.273***	-0.176	-0.408	-0.023	0.105	-0.057	0.032
	(0.489)	(0.498)	(0.087)	(0.091)	(0.277)	(0.357)	(0.072)	(0.083)	(0.054)	(0.029)
UCT	-0.025	0.374	-0.164*	-0.021	0.377	0.613*	-0.031	0.058	-0.063	0.019
	(0.377)	(0.506)	(0.087)	(0.201)	(0.372)	(0.326)	(0.068)	(0.097)	(0.097)	(0.025)
Within CCT EA Control	-0.176	0.002	-0.104***	-0.006	0.065	0.108	-0.037	0.038	0.001	0.007
	(0.195)	(0.523)	(0.035)	(0.060)	(0.175)	(0.243)	(0.032)	(0.036)	(0.044)	(0.018)
Within UCT EA Control	0.333	-0.566	-0.054	-0.107	0.329	0.659*	0.082	-0.110	0.0965*	-0.062
	(0.631)	(0.482)	(0.127)	(0.152)	(0.472)	(0.385)	(0.103)	(0.157)	(0.050)	(0.054)
EA saturation, CCT Treatment	0.873	2.068***	-0.082	0.314***	-0.098	0.248	-0.025	-0.141	0.068	-0.037
	(0.556)	(0.623)	(0.117)	(0.113)	(0.306)	(0.385)	(0.074)	(0.109)	(0.068)	(0.032)
EA saturation, UCT treatment	0.348	0.123	0.124	-0.016	-0.340	-0.370	0.044	-0.010	0.066	-0.018
	(0.425)	(0.524)	(0.121)	(0.219)	(0.437)	(0.451)	(0.081)	(0.119)	(0.126)	(0.029)
EA saturation, Within-CCT Control	0.074	0.263	0.039	0.068	0.169	-0.571	0.167**	0.065	0.033	-0.002
	(0.404)	(1.086)	(0.089)	(0.146)	(0.361)	(0.566)	(0.071)	(0.090)	(0.101)	(0.041)
EA saturation, Within-UCT Control	-1.061	2.099**	-0.045	0.354	-0.696	-1.148	-0.147	0.249	-0.170	0.147
	(1.130)	(0.809)	(0.325)	(0.311)	(0.979)	(0.781)	(0.181)	(0.250)	(0.132)	(0.155)
Mean in Pure Control:	6.432	0.581	0.343	0.616	15.252	7.479	0.842	0.705	0.797	0.023
Observations	2,651	2,651	2,651	2,651	2,653	2,652	2,653	2,648	2,653	2,652

Regressions are OLS using Round 1 data with robust standard errors clustered at the EA level. All regressions are weighted with both sampling and saturation weights to make the results representative of the target population in the study Eas.

TABLE A2. Robustness check using cross-EA variation in treatment intensity

Dependent Variable:	Terms Enrolled		Average Test Score		Ever Married		Ever Pregnant	
	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)
CCT	0.119	0.126	0.022	0.008	0.000	-0.023	0.040	-0.011
	(0.0431)***	(0.085)	(0.00896)**	(0.015)	(0.023)	(0.044)	(0.026)	(0.041)
UCT	0.059	0.052	0.005	-0.019	-0.064	-0.090	-0.057	-0.114
	(0.050)	(0.111)	(0.013)	(0.018)	(0.0269)**	(0.0484)*	(0.0240)**	(0.0508)**
Within CCT EA Control	0.013	0.016	0.021	0.023	0.010	0.011	0.008	0.008
	(0.047)	(0.047)	(0.014)	(0.0134)*	(0.023)	(0.023)	(0.026)	(0.025)
Within UCT EA Control	-0.100	-0.095	-0.020	-0.015	0.000	0.002	-0.021	-0.020
	(0.074)	(0.077)	(0.023)	(0.023)	(0.034)	(0.036)	(0.029)	(0.031)
# of treated EAs within 3 km	-0.021	-0.020	-0.005	-0.002	0.005	0.005	0.004	0.003
	(0.018)	(0.020)	(0.005)	(0.006)	(0.009)	(0.012)	(0.009)	(0.011)
# of treated EAs between 3 & 6 km	0.010	0.019	0.001	0.006	-0.004	-0.002	-0.005	-0.003
	(0.013)	(0.016)	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)	(0.008)
# of total EAs within 3 km	0.012	0.011	0.006	0.004	-0.003	-0.004	0.001	0.002
	(0.012)	(0.013)	(0.00281)**	(0.004)	(0.006)	(0.007)	(0.006)	(0.007)
# of total EAs between 3 & 6 km	-0.004	-0.008	-0.002	-0.004	0.000	-0.001	0.004	0.002
	(0.007)	(0.008)	(0.002)	(0.00216)*	(0.003)	(0.004)	(0.003)	(0.004)
Treated individual * # of treated EAs within 3 kilometers		0.003		0.005		-0.004		-0.007
		(0.021)		(0.004)		(0.011)		(0.012)
Treated individual * # of treated EAs between 3 and 6 kilometers		0.001		-0.006		0.013		0.009
		(0.040)		(0.008)		(0.022)		(0.023)
Treated individual * # of total EAs within 3 kilometers		-0.029		-0.014		-0.007		-0.004
		(0.026)		(0.00467)***		(0.015)		(0.015)
Treated individual * # of total EAs between3 and 6 kilometers		0.012		0.007		0.004		0.007
		(0.014)		(0.00276)**		(0.008)		(0.007)
Observations	2,579	2,579	2,612	2,612	2,649	2,649	2,650	2,650
R-squared	0.098	0.098	0.418	0.42	0.144	0.144	0.199	0.2

Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted with both sampling and saturation weights 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 percent (***), 95 percent (**), and 90 percent (*) confidence.

```
%PROGRAM TO ESTIMATE POWER OF A RANDOMIZED SATURATION STUDY
%AUTHOR: Aislinn Bohren
%SUPPLEMENTAL MATERIAL TO "DESIGNING EXPERIMENTS TO MEASURE SPILLOVER
%EFFECTS" By S. Baird, A. Bohren, C. McIntosh, B. Ozler

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%USER INPUT
clear;clc;

n=20; %cluster size
C=5; %number of clusters
tau=0.5; %variance of cluster error
sigma=2; %variance of individual error
alpha=0.05 ; %significance
gamma=.9; %power

pi=[0,1/3,2/3]; %saturation bins
f=[1/3,1/3,1/3]; %distribution over bins

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%CALCULATIONS
varN=tau+sigma;
varCo=(n-1)*tau;
varC=n*tau+sigma;

%Note: eta=E[pi^2] here; so eta in paper is eta-mu^2
x=1;
powerT=zeros(1,length(x));
powerS=zeros(1,length(x));
powerTonly=zeros(1,length(x));
powerSonly=zeros(1,length(x));
powerT1=zeros(1,length(x));
powerT2=zeros(1,length(x));
powerS1=zeros(1,length(x));
powerS2=zeros(1,length(x));
MDSE_T=zeros(1,length(x));
MDSE_S=zeros(1,length(x));
MDSE_TAffine=zeros(1,length(x));
MDSE_SAffine=zeros(1,length(x));
MU=zeros(1,length(x));
ETA=zeros(1,length(x));
etaT=zeros(1,length(x));

for j=1:length(x);
%Calculate distribution statistics
mu_ind=zeros(1,length(pi));
p=zeros(1,length(pi));
eta_ind=zeros(1,length(pi));
rho=zeros(1,length(pi));
c_ind=zeros(1,length(pi));
d_ind=zeros(1,length(pi));

for i=1:length(pi);
mu_ind(i)=pi(i) *f(i);
eta_ind(i)=pi(i)^2 *f(i);
c_ind(i)=pi(i)^3 *f(i);
d_ind(i)=pi(i)^4 *f(i);
p(i)=(1-pi(i))*f(i);
rho(i)=(1-pi(i))^2 * f(i);
end;

mu=sum(mu_ind);
eta=sum(eta_ind);
c=sum(c_ind);
d=sum(d_ind);
psi=0;if pi(1)==0; psi=f(1);end;
MU(j)=mu;
ETA(j)=eta;
```

```
%PROGRAM TO ESTIMATE POWER OF A RANDOMIZED SATURATION STUDY
%AUTHOR: Aislinn Bohren
%SUPPLEMENTAL MATERIAL TO "DESIGNING EXPERIMENTS TO MEASURE SPILLOVER
%EFFECTS" By S. Baird, A. Bohren, C. McIntosh, B. Ozler

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%USER INPUT
clear;clc;

n=20; %cluster size
C=5; %number of clusters
tau=0.5; %variance of cluster error
sigma=2; %variance of individual error
alpha=0.05 ; %significance
gamma=.9; %power

pi=[0,1/3,2/3]; %saturation bins
f=[1/3,1/3,1/3]; %distribution over bins

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%CALCULATIONS
varN=tau+sigma;
varCo=(n-1)*tau;
varC=n*tau+sigma;

%Note: eta=E[pi^2] here; so eta in paper is eta-mu^2
x=1;
powerT=zeros(1,length(x));
powerS=zeros(1,length(x));
powerTonly=zeros(1,length(x));
powerSonly=zeros(1,length(x));
powerT1=zeros(1,length(x));
powerT2=zeros(1,length(x));
powerS1=zeros(1,length(x));
powerS2=zeros(1,length(x));
MDSE_T=zeros(1,length(x));
MDSE_S=zeros(1,length(x));
MDSE_TAffine=zeros(1,length(x));
MDSE_SAffine=zeros(1,length(x));
MU=zeros(1,length(x));
ETA=zeros(1,length(x));
etaT=zeros(1,length(x));

for j=1:length(x);
%Calculate distribution statistics
mu_ind=zeros(1,length(pi));
p=zeros(1,length(pi));
eta_ind=zeros(1,length(pi));
rho=zeros(1,length(pi));
c_ind=zeros(1,length(pi));
d_ind=zeros(1,length(pi));

for i=1:length(pi);
mu_ind(i)=pi(i) *f(i);
eta_ind(i)=pi(i)^2 *f(i);
c_ind(i)=pi(i)^3 *f(i);
d_ind(i)=pi(i)^4 *f(i);
p(i)=(1-pi(i))*f(i);
rho(i)=(1-pi(i))^2 * f(i);
end;

mu=sum(mu_ind);
eta=sum(eta_ind);
c=sum(c_ind);
d=sum(d_ind);
psi=0;if pi(1)==0; psi=f(1);end;
MU(j)=mu;
ETA(j)=eta;
```