Identification and Estimation of Spillover Effects in Randomized Experiments[*]

Gonzalo Vazquez-Bare[†]

June 15, 2021

Abstract

I study identification, estimation and inference for spillover effects in experiments where units' outcomes may depend on the treatment assignments of other units within a group. I show that the commonly-used reduced-form linear-in-means (RF-LIM) regression identifies a weighted sum of spillover effects with some negative weights, and that the difference in means between treated and controls identifies a combination of direct and spillover effects entering with different signs. I propose nonparametric estimators for average direct and spillover effects that overcome these issues and are consistent and asymptotically normal under a precise relationship between the number of parameters of interest, the total sample size and the treatment assignment mechanism. These findings are illustrated using data from a conditional cash transfer program and with simulations. The empirical results reveal the potential pitfalls of failing to flexibly account for spillover effects in policy evaluation: the estimated difference in means and the RF-LIM coefficients are all close to zero and statistically insignificant, whereas the nonparametric estimators I propose reveal large, nonlinear and significant spillover effects. (*JEL* C10, C13, C14, C90)

[†]Department of Economics, University of California, Santa Barbara. gvazquez@econ.ucsb.edu.

1   Introduction

Spillover effects, which occur when an agent's actions or behaviors indirectly affect other agents' outcomes through peer effects, social interactions or externalities, are ubiquitous in economics and social sciences. A thorough account of spillover effects is crucial to assess the causal impact of policies and programs (Abadie and Cattaneo, 2018; Athey and Imbens, 2017). However, the literature is still evolving in this area, and most of the available methods for analyzing treatment effects either assume no spillovers or allow for them in restrictive ways, sometimes without a precise definition of the parameters of interest or the conditions required to recover them.

This paper studies identification, estimation and inference for average direct and spillover effects in randomized controlled trials, and offers three main contributions. First, I provide conditions for nonparametric identification of causal parameters when the true spillovers structure is possibly unknown. To this end, Section 2 sets up a potential-outcomes framework that nests several models commonly used to analyze spillovers. Under the assumption that interference can occur within (but not between) the groups in which units are clustered, I define a rich set of direct and spillover treatment effects based on a function, the *treatment rule*, that maps peers' treatment assignments and outcomes. Lemma 1 in Section 3 links average potential outcomes to averages of observed variables when the posited treatment rule is possibly misspecified.

The second contribution of this paper is to characterize the difference in means between treated and controls and the coefficients from a reduced-form linear-in-means (RF-LIM) regression, two of the most commonly analyzed estimands when analyzing RCTs and spillover effects in general. Theorem 1 shows that, in the presence of spillovers, the difference in means between treated and controls combines the direct effect of the treatment and the difference in spillover effects for treated and untreated units, and thus the sign of the difference in means is undetermined even when the signs of all direct and spillover effects are known. On the other hand, Theorem 2 shows that a RF-LIM regression recovers a linear combination of spillover effects for different numbers of treated peers where the weights sum to zero, and hence some weights are necessarily negative. As a result, the coefficients from a RF-LIM regression can be zero even when all the spillover effects are non-zero. I then provide sufficient conditions under which the difference in means and the RF-LIM coefficients have a causal interpretation, that is, when they can be written as proper weighted averages of direct and/or spillover effects. I also propose a regression-based pooling strategy that is robust to nonlinearities and heterogeneity in spillover effects.

Finally, the third contribution is to analyze nonparametric estimation and inference for spillover effects. In the presence of spillovers, estimation faces two main challenges: the number of treatment effects to estimate can be large, and the probability of observing units under different treatment assignments can be small. Section 4 provides general conditions for uniform consistency and asymptotic normality of the estimators of interest in a double-array

asymptotic framework where both the number of groups and the number of parameters are allowed to grow with the sample size. This approach highlights the role that the number of parameters and the assignment mechanism play on the asymptotic properties of nonparametric estimators. More precisely, consistency and asymptotic normality are shown under two main conditions that are formalized in the paper: (i) the number of treatment effects should not be "too large" with respect to the sample size, and (ii) the probability of each treatment assignment should not be "too small". These two requirements are directly linked to modeling assumptions on the potential outcomes, the choice of the set of parameters of interest and the treatment assignment mechanism. As an alternative approach to inference based on the normal approximation, the wild bootstrap is shown to be consistent, and simulation evidence suggests that it can yield better performance compared to the Gaussian approximation for moderately large groups in some cases.

The results in this paper are illustrated in a simulation study and using data from a randomized conditional cash transfer. The empirical results clearly highlight the pitfalls of failing to flexibly account for spillovers in policy evaluation: the estimated difference in means and the estimated RF-LIM coefficients are all close to zero and statistically insignificant, whereas the nonparametric estimators I propose reveal large, nonlinear and significant spillover effects.

This paper is related to a longstanding literature on peer effects and social interactions. A large strand of this literature has focused on identification of social interaction effects in parametric models. The most commonly analyzed specification is the linear-in-means (LIM) model, where a unit's outcome is modeled as a linear function of own characteristics, peers' average characteristics and peers' average outcomes (see e.g. Blume, Brock, Durlauf, and Jayaraman, 2015; Kline and Tamer, 2019, for recent reviews). Since Manski (1993)'s critique of LIM models, several strategies have been put forward to identify structural parameters (see e.g. Lee, 2007; Bramoullé, Djebbari, and Fortin, 2009; Davezies, D'Haultfoeuille, and Fougère, 2009; De Giorgi, Pellizzari, and Redaelli, 2010). All these strategies rely on linearity of spillover effects. In this paper, I consider an alternative approach that focuses on reduced-form casual parameters from a potential-outcomes perspective. Within this setup, I show that (reduced-form) response functions can be identified and estimated nonparametrically. While I do not consider identification of structural parameters in this paper, reduced-form parameters are inherently relevant, as they represent the causal effect of changing the peers' covariate values, which is generally more easily manipulable for a policy maker than peers' outcomes (Goldsmith-Pinkham and Imbens, 2013; Manski, 2013a). This is particularly true in my setup, where the covariate of interest is a treatment that is assigned by the policy maker. Furthermore, identification of reduced-form parameters can be thought of as a necessary ingredient for identifying structural models.

On the opposite end of the spectrum, Manski (2013b) and Lazzati (2015) study nonparametric partial identification of response functions under different restrictions on the structural model, the response functions and the structure of social interactions. My paper complements

this important strand of the literature by considering a specific network structure in which spillovers are limited to non-overlapping groups, where the within-group spillovers structure is left unrestricted. This specific structure of social interactions allows me to obtain point (as opposed to partial) identification of causal effects without knowing the true mapping between treatment assignments an potential outcomes in a setting with wide empirical applicability. Furthermore, by focusing on this specific network structure I can analyze the effect of misspecifying the within-group spillovers structure, as discussed in Section 3.

Another body of research has analyzed causal inference under interference from a design-based perspective in which potential outcomes are fixed and all randomness is due to the (known) treatment assignment mechanism (see Tchetgen Tchetgen and VanderWeele, 2012; Ogburn and VanderWeele, 2014; Halloran and Hudgens, 2016, for reviews). Generally, this literature focuses on two aggregate measures of treatment effects, the average direct effect and the average indirect effect, that average over peers' assignments under specific treatment assignment mechanisms such as completely randomized designs (Sobel, 2006) or two-stage randomization (Hudgens and Halloran, 2008, and subsequent studies). In a related paper, Athey, Eckles, and Imbens (2018) derive a procedure to calculate finite-sample randomization-based p-values to test for the presence of spillover effects. My paper complements this literature in several ways. First, I focus on identifying and estimating the entire vector of spillover effects, which can be seen either as the main object of interest, or as an ingredient to construct aggregate summary measures of spillovers (see also Remark 2). Second, my identification results allow for general treatment assignment mechanisms, of which two-stage randomization is a particular case. Third, I consider large-sample inference from a super-population perspective that allows me to disentangle the roles of the number of groups, the number parameters of interest and the treatment assignment mechanism in the performance of estimators and inferential procedures.

Finally, Moffit (2001), Duflo and Saez (2003), Hirano and Hahn (2010) and more recently Baird, Bohren, McIntosh, and Özler (2018) analyze the design of partial population experiments, where spillovers are estimated by exposing experimental units to different proportions of treated peers (or "saturations"). My paper complements this strand of the literature by analyzing identification under general experimental designs and by providing a formal treatment of the effect of the experimental design on inference, which formalizes the advantages of partial population designs. This fact is illustrated in Section 5 and discussed in more detail in Section A2 of the supplemental appendix.

The remainder of the paper is organized as follows. Section 2 describes the setup and defines the parameters of interest. Section 3 provides the main identification results. Section 4 analyzes estimation and inference. Section 5 provides a simulation study, and Section 6 contains the empirical application. Section 7 concludes. The proofs, together with additional results and discussions, are provided in the supplemental appendix.

2   Setup

As a motivating example, consider a program in which parents in low-income households receive a cash transfer from the government as long as their children are enrolled in school and reach a required level of attendance. Suppose that this conditional cash transfer program is evaluated using a randomized pilot in which children are randomly selected to participate. There are several reasons to expect within-household spillovers from this program. On the one hand, the cash transfer may alleviate financial constraints that were preventing the parents from sending their children to school on a regular basis. The program could also help raise awareness on the importance of school attendance, encouraging parents to worry more about sending their children to school. In both these cases, untreated children may indirectly benefit from the program when they have a treated sibling. On the other hand, the program could create incentives for the parents to reallocate resources towards their treated children and away from their untreated siblings, decreasing school attendance for the latter. In all cases, ignoring spillover effects can severely underestimate the costs or the benefits of this policy.

Moreover, these alternative scenarios have drastically different implications on how to assign the program when scaling it up. In the first two situations, treating one child per household can be a cost-effective way to assign the treatment, whereas in the second case, treating all the children in a household can be more beneficial. An accurate assessment of spillovers is therefore crucial for the analysis and designs of public policies.

2.1   Notation and parameters of interest

In light of the motivating example, consider a random sample of groups indexed by $g = 1, \ldots, G$, each with $n_g + 1$ units, so that each unit i in group g has n_g neighbors or peers and $1 \leq n_g < \infty$. I assume group membership is observable. Units in each group are assigned a binary treatment, and a unit's potential outcomes, defined below, can depend on the assignment of all other units in the same group. Using the terminology of Ogburn and VanderWeele (2014), this phenomenon is known as *direct interference*. Interference is assumed to occur between units in the same group, but not between units in different groups.

The individual treatment assignment of unit i in group g is denoted by D_{ig}, taking values $d \in \{0, 1\}$, and the vector of treatment assignments in each group is given by $\mathbf{D}_g = (D_{1g}, \ldots, D_{n_g+1g})$. For each unit i, D_{jig} is the treatment indicator corresponding to unit i's j-th neighbor, collected in the vector $\mathbf{D}_{(i)g} = (D_{1ig}, D_{2ig}, \ldots, D_{n_gig})$. This vector takes values $\mathbf{d}_g = (d_1, d_2, \ldots, d_{n_g}) \in \mathcal{D} \subseteq \{0, 1\}^{n_g}$.

A key element in this setup will be a function $h_0(\cdot)$ that summarizes how the vector $\mathbf{d}_g$ enters the potential outcome. More precisely, define a function or *treatment rule*:

$$h_0 : \mathcal{D} \to \mathcal{H}_0$$

4

that maps $\mathbf{d}_g$ into some value $h_0(\mathbf{d}_g)$ of the same or smaller dimension, so that $\dim(\mathcal{H}_0) \leq \dim(\mathcal{D})$. Following Manski (2013b)'s terminology, for $h_0(\mathbf{d}_g) = \mathbf{h}_0$, I will refer to the tuple $(d, \mathbf{h}_0)$ as the *effective treatment assignment*, an element in the set $\{0,1\} \times \mathcal{H}_0$. The potential outcome for unit i in group g is denoted by the random variable $Y_{ig}(d, \mathbf{h}_0)$ where $\mathbf{h}_0 = h_0(\mathbf{d}_g) \in \mathcal{H}^0$.

Example 1 (SUTVA) If $h_0(\cdot)$ is a constant function, the vector of peers' assignments is ignored and the set of effective treatment assignments becomes $\{0,1\}$, so the potential outcomes do not depend on peers' assignments. In this case the only effective treatment assignments are $D_{ig} = 1$ and $D_{ig} = 0$ (treated and control). This assumption is often known as the stable unit treatment value assumption or SUTVA (Imbens and Rubin, 2015). $\square$

Example 2 (Exchangeability) When potential outcomes depend on how many peers, but not which ones, are treated, peers are said to be *exchangeable*. Exchangeability can be modeled by setting $h_0(\mathbf{d}_g) = \mathbf{1}_g' \mathbf{d}_g$ so $h_0(\cdot)$ summarizes $\mathbf{d}_g$ through the sum of its elements. The set of effective treatment assignments in this case is given by $\{(d, s) : d = 0, 1, s = 0, 1, \ldots, n_g\}$. Exchangeability may be a natural starting point when there is no clear way (or not enough information) to assign identities to peers. $\square$

Example 3 (Stratified exchangeability) Exchangeability may also be imposed by sub-groups. For instance, the vector of assignments may be summarized by the number of male and female treated peers separately. In settings where units are geographically located, peers are commonly assumed to be exchangeable within groups defined by distance such as within one block, between one and two blocks, etc, or by different distance radiuses (e.g. within 100 meters, between 100 and 200 meters and so on). $\square$

Example 4 (Reference groups) When each unit interacts only with a strict subset of her peers, we can define for example $h_0(\cdot) : \{0,1\}^{n_g} \to \{0,1\}^{k_g}$ where $k_g < n_g$. For instance, under the assumption that each unit interacts with her two closest neighbors, $h_0(\mathbf{d}_g) = (d_1, d_2)$ so that $k_g = 2$. The subset of peers with which each unit interacts is known as the *reference group* (Manski, 2013b). $\square$

Example 5 (Non-exchangeable peers) The case in which $h_0(\mathbf{d}_g) = \mathbf{d}_g$ does not provide any dimensionality reduction, as the only restriction it imposes is the existence of a known ordering between peers. This ordering is required to determine who is unit i's nearest neighbor, second nearest neighbor and so on. Such an ordering can be based for example on geographic distance, a spatial weights matrix as used in spatial econometrics, frequency of interaction in a social network, etc. $\square$

In what follows, $\mathbf{0}_g$ and $\mathbf{1}_g$ will denote n_g-dimensional vectors of zeros and ones, respectively. Throughout the paper, I will assume that all the required moments of the potential outcomes are bounded. Unit-level direct effects are defined as differences in potential

outcomes switching own treatment assignment for a fixed peer assignment $\mathbf{h}_0$, $Y_{ig}(1, \mathbf{h}_0) - Y_{ig}(0, \mathbf{h}_0)$. Unit-level spillover effects are defined as differences in potential outcomes switching peer assignments for a fixed own assignment d, $Y_{ig}(d, \mathbf{h}_0) - Y_{ig}(d, \tilde{\mathbf{h}}_0)$.

Given a vector of observed assignments $(D_{ig}, \mathbf{D}_{(i)g})$, the observed outcome is given by $Y_{ig}(D_{ig}, h_0(\mathbf{D}_{(i)g}))$ and can be rewritten as:

$$Y_{ig} = \sum_{d \in \{0,1\}} \sum_{\mathbf{h}_0 \in \mathcal{H}^0} Y_{ig}(d, \mathbf{h}_0) \mathbb{1}(D_{ig} = d) \mathbb{1}(h_0(\mathbf{D}_{(i)g}) = \mathbf{h}_0).$$

To fix ideas, consider a household with three children, $n_g + 1 = 3$. In this household, each kid has two siblings, with assignments d_1 and d_2, so $\mathbf{d}_g = (d_1, d_2)$. If the true treatment rule $h_0(\cdot)$ is the identity function, the potential outcome has the form $Y_{ig}(d, d_1, d_2)$ and hence each unit can have up to $2^{(n_g+1)} = 8$ different potential outcomes. In this case, $Y_{ig}(1, 0, 0) - Y_{ig}(0, 0, 0)$ is the effect of the treatment when both of unit i's siblings are untreated, $Y_{ig}(0, 1, 0) - Y_{ig}(0, 0, 0)$ is the spillover effect on unit i of treating unit i's first sibling, and so on. The average effect of assignment (d, d_1, d_2) compared to $(\tilde{d}, \tilde{d}_1, \tilde{d}_2)$ is thus given by $\mathbb{E}[Y_{ig}(d, d_1, d_2)] - \mathbb{E}[Y_{ig}(\tilde{d}, \tilde{d}_1, \tilde{d}_2)]$. On the other hand, under an exchangeable treatment rule, the potential outcome can be written as $Y_{ig}(d, s)$ where $s = 0, 1, 2$ is the number of treated siblings. The total number of different potential outcomes is $2(n_g + 1) = 6$, so exchangeability reduces the dimensionality of the effective assignments set from exponential to linear in group size.

I will assume perfect compliance, which means that all units receive the treatment they are assigned to. I analyze the case of imperfect compliance in Vazquez-Bare (2021). The data come from an infinite population of groups for which the researcher observed the outcomes at the unit level and the vector of treatment assignments in each group. Furthermore, by virtue of random assignment, potential outcomes are independent of treatment assignment. I formalize these features as follows.

Assumption 1 (Sampling and random assignment) *Let $\mathbf{y}_{ig} = (Y_{ig}(d, \mathbf{h}_0))_{(d, \mathbf{h}_0) \in \{0,1\} \times \mathcal{H}_0}$ be the vector of potential outcomes for each unit i in group g and $\mathbf{y}_g = (\mathbf{y}'_{1g}, \ldots, \mathbf{y}'_{n_g+1, g})'$.*

(a) *The vectors $(\mathbf{y}'_1, \mathbf{D}'_1)', \ldots, (\mathbf{y}'_G, \mathbf{D}'_G)'$ are sampled independently from an infinite population.*

(b) *For any g and l such that $n_g = n_l$, $(\mathbf{y}'_g, \mathbf{D}'_g)'$ and $(\mathbf{y}'_l, \mathbf{D}'_l)'$ are identically distributed.*

(c) *For each g, the elements in $\mathbf{y}_g$ are identically distributed.*

(d) *For each g, $\mathbf{y}_g \perp\!\!\!\perp \mathbf{D}_g$.*

Parts (a) and (b) in Assumption 1 state that we observe a random sample of independent groups, and all groups with the same size are identically distributed. Part (c) indicates that units are identically distributed within each group, so that, for instance, $\mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]$ is not indexed by i (or g, given part (b)), and the same holds for other moments of the potential outcomes. Note that this condition does not prevent average potential outcomes

(or other moments) to differ conditional on covariates. For example, if X_{ig} denotes gender, the setup allows for $\mathbb{E}[Y_{ig}(d, \mathbf{h}_0)|X_{ig} = male] \neq \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)|X_{ig} = female]$. Additionally, the identification results in the paper can adapted to the case of non-identical distributions within group by switching focus from $\mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]$ to $\sum_{i=1}^{n_g+1} \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]/(n_g + 1)$. Finally, part (d) states that the treatment is randomly assigned and hence the vector of treatment indicators is independent of potential outcomes.

In practice, the true $h_0(\cdot)$ is usually unknown, and the researcher needs to posit a candidate $h(\cdot)$ that may or may not coincide with $h_0(\cdot)$. Given the lack of knowledge on the true assignment, a function $h(\cdot)$ that imposes fewer restrictions on the potential outcomes has a lower risk of misspecification. To formalize this idea, I introduce the following definition.

Definition 1 (Coarseness) *Given two treatment rules $h(\cdot) : \mathcal{D} \to \mathcal{H}$ and $\tilde{h}(\cdot) : \mathcal{D} \to \tilde{\mathcal{H}}_g$, we say $h(\cdot)$ is coarser than $\tilde{h}(\cdot)$ if there exists another function $f(\cdot) : \tilde{\mathcal{H}}_g \to \mathcal{H}$ such that $h(\mathbf{d}_g) = f \circ \tilde{h}(\mathbf{d}_g)$ for all $\mathbf{d}_g \in \mathcal{D}$.*

Intuitively, this means that $h(\cdot)$ gives a "cruder" summary of $\mathbf{d}_g$ (i.e. it discards more information) than $\tilde{h}(\cdot)$. In other words, a coarser function imposes more restrictions on the potential outcomes. For example, the exchangeable assignment $h(\mathbf{d}_g) = \mathbf{1}'_g \mathbf{d}_g$ is coarser than the identity function $\tilde{h}(\mathbf{d}_g) = \mathbf{d}_g$, and the reference group assignment $h(\mathbf{d}_g) = (d_1, d_2)$ is coarser than $\tilde{h}(\mathbf{d}_g) = (d_1, d_2, d_3, d_4)$.

The next section addresses identification of average potential outcomes when the true treatment rule $h_0(\cdot)$ is possibly unknown.

3   Identification

In what follows, let $\mathbf{H}_{ig} = h(\mathbf{D}_{(i)g})$ be the observed value of the chosen treatment rule, and let $\mathbf{H}_{ig}^0 = h_0(\mathbf{D}_{(i)g})$. The following result links observed outcomes, potential outcomes and effective treatment assignments, and will be used for the upcoming theorems.

Lemma 1 (Nonparametric Identification) *Suppose Assumption 1 holds and let $h_0(\cdot) : \mathcal{D} \to \mathcal{H}_0$ be the true treatment rule. Given a treatment rule $h(\cdot) : \mathcal{D} \to \mathcal{H}$, for any pair $(d, \mathbf{h}) \in \{0, 1\} \times \mathcal{H}$ such that $\mathbb{P}[D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] > 0$ and for any measurable function $m(\cdot)$,*

$$\mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[m(Y_{ig}(d, \mathbf{h}_0))]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}].$$

In particular, if $h_0(\cdot)$ is coarser than $h(\cdot)$, then

$$\mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \mathbb{E}[m(Y_{ig}(d, h_0(\mathbf{h})))].$$

In what follows, unless explicitly stated, I will let $m(\cdot)$ be the identity function to reduce notation. More generally, different choices of the function $m(\cdot)$ lead to different estimands of interest. For example, setting $m(\cdot) = \mathbb{1}(\cdot \leq y)$ for some $y \in \mathbb{R}$ yields $\mathbb{E}[m(Y_{ig})|D_{ig} =$

$d, \mathbf{H}_{ig} = \mathbf{h}] = F_Y(y|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h})$ where $F_Y(\cdot|\cdot)$ is the conditional cdf of Y_{ig}. This choice of $m(\cdot)$ can be used to identify the distribution of potential outcomes.

Lemma 1 shows that the average observed outcome among units facing $D_{ig} = d$ and $\mathbf{H}_{ig} = \mathbf{h}$ averages the potential outcomes over all the assignments $\mathbf{h}_0$ that are consistent with $(D_{ig}, \mathbf{H}_{ig}) = (d, \mathbf{h})$, as long as the probability of $(d, \mathbf{h})$ is not zero.

To illustrate Lemma 1, consider the previous example with three units and where $h_0(\cdot)$ is the identity function so the potential outcome has the form $Y_{ig}(d, d_1, d_2)$. Suppose we posit an exchangeable treatment rule $h(\mathbf{d}_g) = \mathbf{1}'_g \mathbf{d}_g$ and thus $\mathbf{H}_{ig} = S_{ig} = \sum_{j \neq i} D_{jg}$ which is a scalar counting how many of unit i's peers are treated. By Lemma 1, if $\mathbb{P}[D_{ig} = 0, S_{ig} = 1] > 0$, $\mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 1]$ equals a weighted average of $\mathbb{E}[Y_{ig}(0,1,0)]$ and $\mathbb{E}[Y_{ig}(0,0,1)]$, with weights given by the conditional probabilities of these different assignments.

In general, $(D_{ig}, \mathbf{H}_{ig}) = (d, \mathbf{h})$ may be consistent with many different effective assignments $\mathbf{h}_0$. When $h_0(\cdot)$ is coarser than $h(\cdot)$, however, the value of $\mathbf{h_0}$ is uniquely determined. In such cases, the second part of Lemma 1 shows that $\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}]$ identifies the value of the average potential outcome consistent with that assignment. For example, suppose that the true $h_0(\cdot)$ is exchangeable, so that outcomes have the form $Y_{ig}(d, s)$ with $s = 0, 1, 2$, and suppose we posit $h(\mathbf{d}_g) = \mathbf{d}_g = (d_1, d_2)$. Setting $\mathbf{H}_{ig} = (1,1)$ implies that the sum of treated peers is equal to 2, and therefore $\mathbb{E}[Y_{ig}|D_{ig} = 0, \mathbf{H}_{ig} = (1,1)] = \mathbb{E}[Y_{ig}(0,2)]$. In particular, this result implies that if the function $h(\cdot)$ is correctly specified, $\mathbb{E}[Y_{ig}|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \mathbb{E}[Y_{ig}(d, \mathbf{h})]$.

Remark 1 (Implications for $h_0(\cdot)$) When $h_0(\cdot)$ is coarser than $h(\cdot)$, Lemma 1 implies restrictions on the shape of the true treatment rule $h_0(\cdot)$. More precisely, if there exists an $m(\cdot)$ and a pair $(\mathbf{h}, \tilde{\mathbf{h}})$ such that $\mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] \neq \mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \tilde{\mathbf{h}}]$, then it follows that $h_0(\mathbf{h}) \neq h_0(\tilde{\mathbf{h}})$.[1] For example, suppose each unit has two peers and that $h(\mathbf{d}_g) = \mathbf{d_g} = (d_1, d_2)$. In this case, if $\mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = (1,0)] \neq \mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = (0,1)]$, then $h_0(1,0) \neq h_0(0,1)$ which rules out, for instance, an exchangeable treatment rule $h_0(d_1, d_2) = d_1 + d_2$ and an "interaction" treatment rule $h_0(d_1, d_2) = d_1 \cdot d_2$. On the other hand, finding that $\mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] = \mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \tilde{\mathbf{h}}]$ for some $m(\cdot)$ does not imply that $h_0(\mathbf{h}) = h_0(\tilde{\mathbf{h}})$, since the equality between moments could fail for a different choice of $m(\cdot)$. $\square$

Remark 2 (Pooled estimands) Coarse treatment rules can be used not only as a modeling assumption on potential outcomes but also as summary measures of average potential outcomes and treatment effects. This may be particularly useful when $h_0(\cdot)$ is suspected to be high-dimensional. For instance, setting $h(\cdot)$ equal to a constant function, which ignores $\mathbf{d}_g$, averages over all possible peers' assignments: $\mathbb{E}[Y_{ig}|D_{ig} = d] = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}^0_{ig} = \mathbf{h}_0|D_{ig} = d]$. Notice that $\mathbb{E}[Y_{ig}|D_{ig} = 1] - \mathbb{E}[Y_{ig}|D_{ig} = 0] = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(1, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}^0_{ig} =$

---

[1]To see this, note that $\mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \mathbf{h}] \neq \mathbb{E}[m(Y_{ig})|D_{ig} = d, \mathbf{H}_{ig} = \tilde{\mathbf{h}}]$ implies $\mathbb{E}[m(Y_{ig}(d, h_0(\mathbf{h})))] \neq \mathbb{E}[m(Y_{ig}(d, h_0(\tilde{\mathbf{h}})))]$ by Lemma 1. If $h_0(\mathbf{h}) = h_0(\tilde{\mathbf{h}})$, then $\mathbb{E}[m(Y_{ig}(d, h_0(\mathbf{h})))] \neq \mathbb{E}[m(Y_{ig}(d, h_0(\mathbf{h})))]$ which gives a contradiction, and thus $h_0(\mathbf{h}) \neq h_0(\tilde{\mathbf{h}})$.

$\mathbf{h}_0|D_{ig} = 1] - \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(0, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = 0]$, which is the the super-population analog of the direct average causal effect defined by Hudgens and Halloran (2008). Alternatively, let $s = \mathbf{1}_g'\mathbf{d}_g$, and define $h(\mathbf{d}_g) = \mathbb{1}(s > 0)$ which equals one if there is at least one treated peer. Let $S_{ig} = \sum_{j \neq i} D_{jg}$ be the observed number of treated peers for unit i. Then, by Lemma 1, $\mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig} > 0] = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(d, \mathbf{h}_0)]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = d, S_{ig} > 0]$. Consider the difference between untreated units with at least one treated peer and untreated units with no treated peers:

$$\Delta = \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} > 0] - \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 0].$$

Then, given that $S_{ig} = 0$ implies that $\mathbf{D}_{(i)g} = \mathbf{0}_g$, we have that:

$$\Delta = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(0, \mathbf{h}_0) - Y_{ig}(0, \mathbf{0})]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = 0, S_{ig} > 0]$$

where $\mathbf{0} = h_0(\mathbf{0}_g)$. Thus, Δ recovers a weighted average of spillover effects on untreated units weighted by the probabilities of the different assignments, which is analogous to the average indirect causal effect of Hudgens and Halloran (2008) but for a general assignment mechanism. A natural generalization of this idea is to split S_{ig} into categories such as $S_{ig} = 0$, $1 \leq S_{ig} \leq k$, $k + 1 \leq S_{ig} \leq n_g$ and so on. Section 6 illustrates how to estimate these pooled parameters using linear regressions. $\square$

Remark 3 (Partial population experiments) A popular design when analyzing spillover effects is the partial population design (Moffit, 2001; Duflo and Saez, 2003; Hirano and Hahn, 2010; Baird, Bohren, McIntosh, and Özler, 2018). In its simplest form, groups are randomly divided into treated and controls based on a binary indicator T_g. Then, within the groups with $T_g = 1$, treatment D_{ig} is randomly assigned at the individual level. In these type of experiments, spillover effects are often estimated as the average difference between control units in treated groups and control units in pure control groups,

$$\Delta_{\mathsf{PP}} = \mathbb{E}[Y_{ig}|D_{ig} = 0, T_g = 1] - \mathbb{E}[Y_{ig}|T_g = 0].$$

Redefining the vector of treatment assignments as $(D_{ig}, \mathbf{D}_{(i)g}, T_g) = (d, \mathbf{d}_g, t)$ and setting $h(\mathbf{d}_g, t) = t$, if $(\mathbf{D}_g, T_g)$ is independent of potential outcomes, Lemma 1 implies that:

$$\Delta_{\mathsf{PP}} = \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(0, \mathbf{h}_0) - Y_{ig}(0, \mathbf{0})]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = 0, T_g = 1]$$

which averages over all the possible number of treated peers that an untreated unit can have in a treated group. The generalization to experiments with more than two categories (see e.g. Crépon, Duflo, Gurgand, Rathelot, and Zamora, 2013) is straightforward. $\square$

3.1 Difference in means

The difference in means estimand $\beta_{\mathsf{D}} = \mathbb{E}[Y_{ig}|D_{ig} = 1] - \mathbb{E}[Y_{ig}|D_{ig} = 0]$, which compares the average observed outcomes between treated and controls, is arguably the most common estimand when analyzing randomized experiments. It is well known that, in the absence of spillovers, β_{D} equals the average treatment effect (ATE) when the treatment is randomly assigned. An estimate for the ATE can be calculated by estimating the model:

$$Y_{ig} = \alpha_{\mathsf{D}} + \beta_{\mathsf{D}} D_{ig} + u_{ig}, \quad \mathbb{E}[u_{ig}] = \mathbb{Cov}(D_{ig}, u_{ig}) = 0. \tag{1}$$

The following results characterizes the difference in means β_{D} in the presence of spillovers. In what follows, let $\mathbf{0} = h_0(\mathbf{0}_g)$.

Theorem 1 (Difference in means) *Under Assumption 1, the coefficient β_{D} from Equation (1) can be written as:*

$$\beta_{\mathsf{D}} = \mathbb{E}[Y_{ig}(1,\mathbf{0}) - Y_{ig}(0,\mathbf{0})] + \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(1,\mathbf{h}_0) - Y_{ig}(1,\mathbf{0})]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = 1]$$
$$- \sum_{\mathbf{h}_0 \in \mathcal{H}_0} \mathbb{E}[Y_{ig}(0,\mathbf{h}_0) - Y_{ig}(0,\mathbf{0})]\mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}_0|D_{ig} = 0].$$

Hence, the difference-in-means estimand equals the direct effect without treated peers $\mathbb{E}[Y_{ig}(1,\mathbf{0}) - Y_{ig}(0,\mathbf{0})]$ plus the difference in weighted averages of spillover effects under treatment and under control. In general, the sign of this difference is undetermined, as it depends on the relative magnitudes of the average spillover effects on treated and controls. Hence, the difference in means can be large or smaller than or equal to the average direct effect $\mathbb{E}[Y_{ig}(1,\mathbf{0}) - Y_{ig}(0,\mathbf{0})]$. In particular, if the spillover effects on treated units are equal to zero and the spillover effects on controls are of the same sign that the direct effects, the difference in means will underestimate the average direct effect without treated peers. In this case, the presence of spillovers pushes the difference in means towards zero, which captures the idea of "contamination" of the control group.

3.2 Reduced-form linear-in-means regression

Equation (1) may give an incomplete assessment of the effect of a program because it completely ignores the presence of spillovers. When trying to explicitly estimate spillover effects, a common strategy is to estimate a reduced-form linear-in-means (RF-LIM) regression, which is given by:

$$Y_{ig} = \alpha_{\ell} + \beta_{\ell} D_{ig} + \gamma_{\ell} \bar{D}_g^{(i)} + \eta_{ig}, \quad \mathbb{E}[\eta_{ig}] = \mathbb{Cov}(D_{ig}, \eta_{ig}) = \mathbb{Cov}(\bar{D}_g^{(i)}, \eta_{ig}) = 0 \tag{2}$$

where

$$\bar{D}_g^{(i)} = \frac{1}{n_g} \sum_{j \neq i} D_{jg}$$

This is a regression of the outcome on a treatment indicator and the proportion of treated peers. In this specification, β_ℓ intends to capture a direct effect whereas γ_ℓ is seen as a measure of spillover effects, since it captures the average change in outcomes in response to a change in the proportion of treated neighbors. While the parameters $(\beta_\ell, \gamma_\ell)$ can be interpreted as coefficients from a linear projection, the following result shows that they do not have a causal interpretation in general.

Theorem 2 (RF-LIM regression) *Under Assumption 1, the coefficients $(\beta_\ell, \gamma_\ell)$ from Equation (2) can be written as:*

$$\beta_\ell = \mathbb{E}[Y_{ig}|D_{ig}=1] - \mathbb{E}[Y_{ig}|D_{ig}=0] - \frac{\gamma_\ell}{n_g}(\mathbb{E}[S_{ig}|D_{ig}=1] - \mathbb{E}[S_{ig}|D_{ig}=0])$$

$$\gamma_\ell = \sum_{s=1}^{n_g} \phi_0(s)(\mathbb{E}[Y_{ig}|D_{ig}=0, S_{ig}=s] - \mathbb{E}[Y_{ig}|D_{ig}=0, S_{ig}=0])$$
$$+ \sum_{s=1}^{n_g} \phi_1(s)(\mathbb{E}[Y_{ig}|D_{ig}=1, S_{ig}=s] - \mathbb{E}[Y_{ig}|D_{ig}=1, S_{ig}=0])$$

where for $d = 0, 1$,

$$\phi_d(s) = \frac{n_g \mathbb{P}[D_{ig}=d]\mathbb{P}[S_{ig}=s|D_{ig}=d]}{\mathbb{P}[D_{ig}=0]\mathbb{V}[S_{ig}|D_{ig}=0] + \mathbb{P}[D_{ig}=1]\mathbb{V}[S_{ig}|D_{ig}=1]} \cdot (s - \mathbb{E}[S_{ig}|D_{ig}=d]).$$

Theorem 2 characterizes the coefficients from the linear projection of Y_{ig} into D_{ig} and $\bar{D}_g^{(i)}$. The coefficient β_ℓ equals the difference in means minus an adjustment factor that depends on γ_ℓ and the relationship between treatment assignments within group. This parameter is similar to the one analyzed in Theorem 1 and hence does not have a direct causal interpretation in general.

The coefficient γ_ℓ equals a linear combination of differences $\mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s] - \mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=0]$ across all values of s with weights $(\phi_0(s), \phi_1(s))_s$. Two factors obscure the causal interpretation of this coefficient. On the one hand, the magnitudes $\mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s]$ implicitly impose an exchangeable treatment rule that may be misspecified in general. The interpretation of these expectations is given in Lemma 1. On the other hand, even if this treatment rule was correctly specified, so that $\mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s] - \mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=0] = \mathbb{E}[Y_{ig}(d,s) - Y_{ig}(d,0)]$, these effects are combined using weights that can be negative. Indeed, it can be seen that $\sum_{s=0}^{n_g} \phi_d(s) = 0$ for $d = 0, 1$ and hence some of the weights are necessarily negative (specifically, the ones corresponding to low values of s). Thus, γ_ℓ may be zero or negative when all the average spillover effects are positive or vice versa.

To illustrate the importance of these issues in an empirical setting, Section 6 shows a case in which the estimates of β_ℓ and γ_ℓ are close to zero and not statistically significant, even when the estimated average spillover effects are all large and statistically significant when estimated nonparametrically.

The following result provides conditions under which the coefficients from a RF-LIM regression have a clear causal interpretation, that is, they can be written as proper weighted averages of direct and/or spillover effects.

Corollary 1 (Correctly-specified RF-LIM regression) *Suppose that, in addition to Assumption 1, the following conditions hold:*

(i) Exchangeability: $Y_{ig}(d, h_0(\mathbf{d}_g)) = Y_{ig}(d, s)$ *where* $s = \mathbf{1}_g' \mathbf{d}_g$,

(ii) Linearity: for each $d = 0, 1$ there is a constant κ_d such that $\mathbb{E}[Y_{ig}(d, s) - Y_{ig}(d, s-1)] = \kappa_d$ for all $s \geq 1$.

Then, the coefficients $(\beta_\ell, \gamma_\ell)$ from Equation (2) are:

$$\beta_\ell = \mathbb{E}[Y_{ig}(1,0) - Y_{ig}(0,0)] + (\kappa_1 - \kappa_0)\{(1-\lambda)\mathbb{E}[S_{ig}|D_{ig} = 1] + \lambda\mathbb{E}[S_{ig}|D_{ig} = 0]\}$$
$$\gamma_\ell = \lambda\mathbb{E}[Y_{ig}(1, n_g) - Y_{ig}(1,0)] + (1-\lambda)\mathbb{E}[Y_{ig}(0, n_g) - Y_{ig}(0,0)]$$

where

$$\lambda = \frac{\mathbb{P}[D_{ig} = 1]\mathbb{V}[S_{ig}|D_{ig} = 1]}{\mathbb{P}[D_{ig} = 1]\mathbb{V}[S_{ig}|D_{ig} = 1] + \mathbb{P}[D_{ig} = 0]\mathbb{V}[S_{ig}|D_{ig} = 0]} \in (0, 1).$$

The above result highlights two restrictions that the RF-LIM regression implicitly imposes on potential outcomes: (i) peers are exchangeable, so potential outcomes only depend on own treatment and the number of treated peers, and (ii) average spillover effects are linear in s so that, for instance, the effect of having two treated peers is twice as large as the effect of having one treated peer. If these conditions hold, γ_ℓ recovers a weighted average of the effects of having all treated peers for treated and untreated units, where the weights are positive and sum to one. Hence, the RF-LIM regression is robust to some heterogeneity in spillover effects $\mathbb{E}[Y_{ig}(d, s) - Y_{ig}(d, 0)]$ both over s and over d, but suffers from potentially severe misspecification when spillover effects are nonlinear.

On the other hand, β_ℓ does not recover a causal effect in general. In the particular case in which $\kappa_1 = \kappa_0$, that is, the average spillover effects are equal for treated and untreated units, β_ℓ equals the average direct effect with no treated peers $\mathbb{E}[Y_{ig}(1,0) - Y_{ig}(0,0)]$.

Remark 4 (Structural LIM models) Structural LIM models include average peers outcomes $\bar{Y}_g^{(i)}$ on the right-hand side of an equation like (2) (Manski, 1993; Kline and Tamer, 2019). Under this specification, in addition to the dependence on peers' treatments, a unit's outcome can be affected by peers' outcomes, a phenomenon known as "endogenous effects" (Manski, 1993) or "interference by contagion" (Ogburn and VanderWeele, 2014). While I do not consider identification of endogenous effects in this paper, Section A1 in the supplemental appendix shows that Equation (2) can be rationalized as the reduced form of the structural LIM model:

$$Y_{ig} = \phi(D_{ig}, \mathbf{D}_{(i)g}) + \gamma\bar{Y}_g^{(i)} + u_{ig}$$

under the assumptions of Corollary 1. $\square$

A straightforward way to make Equation (2) more flexible is to include an interaction term between own treatment indicator and the proportion of treated peers:

$$Y_{ig} = \tilde{\alpha}_\ell + \tilde{\beta}_\ell D_{ig} + \gamma_\ell^0 \bar{D}_g^{(i)}(1 - D_{ig}) + \gamma_\ell^1 \bar{D}_g^{(i)} D_{ig} + \xi_{ig} \tag{3}$$

where $\mathbb{E}[\xi_{ig}] = \mathbb{C}\mathrm{ov}(D_{ig}, \xi_{ig}) = \mathbb{C}\mathrm{ov}(\bar{D}_g^{(i)}(1 - D_{ig}), \xi_{ig}) = \mathbb{C}\mathrm{ov}(\bar{D}_g^{(i)} D_{ig}, \xi_{ig}) = 0$. The following result characterizes the coefficients from this specification.

Theorem 3 (Interacted RF-LIM regression) *Under Assumption 1, the coefficients* $(\tilde{\beta}_\ell, \gamma_\ell^0, \gamma_\ell^1)$ *from Equation (3) can be written as:*

$$\tilde{\beta}_\ell = \mathbb{E}[Y_{ig}|D_{ig} = 1] - \mathbb{E}[Y_{ig}|D_{ig} = 0] - \left(\frac{\gamma_\ell^1}{n_g} \mathbb{E}[S_{ig}|D_{ig} = 1] - \frac{\gamma_\ell^0}{n_g} \mathbb{E}[S_{ig}|D_{ig} = 0] \right)$$

$$\gamma_\ell^0 = \sum_{s=1}^{n_g} \omega_0(s)(\mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = s] - \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 0])$$

$$\gamma_\ell^1 = \sum_{s=1}^{n_g} \omega_1(s)(\mathbb{E}[Y_{ig}|D_{ig} = 1, S_{ig} = s] - \mathbb{E}[Y_{ig}|D_{ig} = 1, S_{ig} = 0])$$

where for $d = 0, 1$,

$$\omega_d(s) = \frac{n_g \mathbb{P}[S_{ig} = s|D_{ig} = d]}{\mathbb{V}[S_{ig}|D_{ig} = d]} \cdot (s - \mathbb{E}[S_{ig}|D_{ig} = d]).$$

According to this theorem, an interacted RF-LIM regression separates the spillover components $\mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = s] - \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 0]$ and $\mathbb{E}[Y_{ig}|D_{ig} = 1, S_{ig} = s] - \mathbb{E}[Y_{ig}|D_{ig} = 1, S_{ig} = 0]$. However, the issue of negative weights remains for each component, since $\sum_{s=0}^{n_g} \omega_d(s) = 0$ for $d = 0, 1$ and hence some of the weights are necessarily negative.

Finally, the following result shows that when peers are exchangeable and spillover effects are linear, the interacted RF-LIM regression can recover all the causal parameters of interest.

Corollary 2 (Correctly-specified interacted RF-LIM regression) *Suppose that, in addition to Assumption 1, the following conditions hold:*

(i) Exchangeability: $Y_{ig}(d, h_0(\mathbf{d}_g)) = Y_{ig}(d, s)$ *where* $s = \mathbf{1}_g' \mathbf{d}_g$,

(ii) Linearity: for each $d = 0, 1$ *there is a constant* κ_d *such that* $\mathbb{E}[Y_{ig}(d, s) - Y_{ig}(d, s - 1)] = \kappa_d$ *for all* $s \geq 1$.

Then, the coefficients $(\tilde{\beta}_\ell, \gamma_\ell^0, \gamma_\ell^1)$ *from Equation (3) can be written as:*

$$\tilde{\beta}_\ell = \mathbb{E}[Y_{ig}(1, 0) - Y_{ig}(0, 0)]$$

$$\gamma_\ell^0 = \mathbb{E}[Y_{ig}(0, n_g) - Y_{ig}(0, 0)]$$

$$\gamma_\ell^1 = \mathbb{E}[Y_{ig}(1, n_g) - Y_{ig}(1, 0)].$$

According to this result, the coefficients from a correctly-specified interacted RF-LIM regression recover the average direct effect without treated peers and the spillover effects of

13

having all peers treated, for treated and untreated units separately. Because of linearity, all the remaining spillover effects can be recovered by appropriately rescaling γ_ℓ^d. For example, the average spillover effect from having one treated peer on an untreated unit is $\mathbb{E}[Y_{ig}(0,1) - Y_{ig}(0,0)] = \gamma_\ell^0/n_g$.

4 Estimation and inference

The previous sections provide conditions under which average direct and spillover effects can be nonparametrically identified by exploiting variation over own and peers' assignments. Because these magnitudes can be written as population averages, it is straightforward to construct their corresponding estimators based on sample cell means. The main challenge for estimation arises when groups are large. A large number of units per group requires estimating a large number of means in each of the cells defined by the treatment assignments. When groups have many units (as in households with many siblings or classrooms with a large number of students), the probability of observing some assignments can be close to zero and the number of observations in each cell can be too small to estimate the average potential outcomes.

For example, suppose the treatment is assigned as an independent coin flip with probability $p = 1/2$. Under this assignment we would expect most groups to have about half its units treated, so when groups have, say, 10 units, 5 of them would be treated on average. The probability of observing groups with zero or all treated units, on the other hand, will be close to zero, and thus the average potential outcomes corresponding to these "tail assignments" will be very hard to estimate.

So far, the analysis has been done taking group size as fixed. When group size is fixed, small cells are a finite sample problem that disappears as the sample grows. To account for this phenomenon asymptotically, in this section I will generalize this setting and consider double-array asymptotics in which group size is allowed to grow with the sample size. The goal is to answer the question of how large groups can be relative to the total sample size to allow for valid estimation and inference. The key issue to obtain consistency and asymptotic normality will be to ensure that the number of observations in all cells grows to infinity as the sample size increases. This setup is not intended to model a population in which groups are effectively infinitely large, but as a statistical device to approximate the distribution of estimators in finite samples when the number of parameters can be "moderately" large, in a sense that will be made more precise in this section. The case with fixed group size is a particular case of this setting.

In this section I will assume that groups are equally sized, so that $n_g = n$. Recall that given a candidate treatment rule $h(\cdot)$ and $\mathbf{h} = h(\mathbf{d}_g)$, the effective treatment assignments are given by $(d, \mathbf{h}_g) \in \{0,1\} \times \mathcal{H}$. As formalized in Assumption 2 below, $h(\cdot)$ is not assumed to equal the true treatment rule, but the true treatment rule $h_0(\cdot)$ has to be coarser than $h(\cdot)$ as specified in Definition 1. To make the notation more compact, I will let $\mathcal{A}_n = \{0,1\} \times \mathcal{H}$ where

the notation makes the dependence of this set on the group size explicit. The cardinality of this set is denoted by $|\mathcal{A}_n|$, which indicates the total number of parameters to be estimated. The observed effective treatment assignments will be $(D_{ig}, \mathbf{H}_{ig}) = \mathbf{A}_{ig}$, taking values $\mathbf{a} \in \mathcal{A}_n$, and $\mu(\mathbf{a}) = \mathbb{E}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}]$.

Because $\mathbf{A}_{ig}$ takes on a finite number of values, all the conditional means can be estimated jointly through the regression:

$$Y_{ig} = \alpha + \sum_{\mathbf{a} \in \mathcal{A}_{n,0}} \beta_{\mathbf{a}} \mathbb{1}(\mathbf{A}_{ig} = \mathbf{a}) + \nu_{ig}, \quad \mathbb{E}[\nu_{ig}|\mathbf{A}_{ig}] = 0 \tag{4}$$

where $\mathcal{A}_{n,0} = \mathcal{A}_n \backslash \{\mathbf{a}_0\}$ and $\mathbf{a}_0$ is the baseline treatment assignment (typically, the assignment in which no unit is treated). Because this regression is fully saturated, by construction $\alpha = \mathbb{E}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}_0]$, $\beta_{\mathbf{a}} = \mathbb{E}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}] - \mathbb{E}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}_0]$. Hence, this regression can be seen as a nonparametric regression as it does not impose any modeling assumptions. Since all the coefficients are linear combinations of conditional means, it suffices to focus on the vector of means $\mu(\mathbf{a})$ to analyze the properties of the coefficient estimators.

Each treatment assignment mechanism determines a distribution $\pi(\cdot)$ over $\mathcal{A}_n$ where $\pi(\mathbf{a}) = \mathbb{P}[\mathbf{A}_{ig} = \mathbf{a}]$ for $\mathbf{a} \in \mathcal{A}_n$. For example, when $\mathcal{A}_n = \{0, 1\}$, if the treatment is assigned independently as a coin flip, $\pi(1) = \mathbb{P}[D_{ig} = 1] = p$ and $\pi(0) = 1 - p$. Under the same assignment, with an exchangeable treatment rule, $\pi(d, s) = \mathbb{P}[D_{ig} = d, S_{ig} = s] = \binom{n}{s} p^{s+d}(1-p)^{n+1-s-d}$. A key issue of this double-array asymptotic setup is that, since the size of the set $\mathcal{A}_n$ can increase with group size, the probabilities $\pi(\mathbf{a})$ can shrink towards zero for some (or all) assignments $\mathbf{a} \in \mathcal{A}_n$. The rate at which these probabilities decrease with the sample size is given by the experimental design. For instance, in the coin flip assignment just described, $\mathbb{P}[D_{ig} = 0, S_{ig} = 0] = (1-p)^{n+1}$ which decreases exponentially with n. Define:

$$\underline{\pi}_n = \min_{\mathbf{a} \in \mathcal{A}_n} \pi(\mathbf{a})$$

which is the probability of the least likely treatment assignment. This probability, together with the total sample size, will determine the number of observations in the smallest assignment cell, that is, the number of observations available to estimate the "hardest" average potential outcome.

Let $\mathbf{A}_g = (\mathbf{A}_{1g}, \ldots, \mathbf{A}_{n_g+1,g})$, $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_G)$, and $\mathbf{Y}_g = (Y_{1g}, Y_{2g}, \ldots Y_{n_g+1,g})'$. I will assume the following.

Assumption 2 (Sampling and design)

  (i) *For $g = 1, \ldots, G$, $(\mathbf{Y}_g', \mathbf{A}_g')$ are iid, and $n_g = n$.*

  (ii) *The true treatment rule $h_0(\cdot)$ is coarser than $h(\cdot)$.*

  (iii) *The potential outcomes are independent across i within groups.*

  (iv) *$|\mathcal{A}_n| = O(G(n+1)\underline{\pi}_n)$, as $G \to \infty$ and $n \to \infty$.*

Part (i) in Assumption 2 states that the researcher has access to a sample of G independent and identically distributed equally-sized groups. When groups have different sizes (for example, households with 3, 4 or 5 siblings), the analysis can be performed separately for each group size. Section A4 of the supplemental appendix further discusses the case of unequally-sized groups. Part (ii) allows the posited treatment rule $h(\cdot)$ to be different from the true treatment rule, but requires it to be flexible enough to break the dependence between Y_{ig} and $\mathbf{A}_{jg}$ conditional on $\mathbf{A}_{ig}$ for $j \neq i$. Part (iii) assumes that potential outcomes are independent within groups, and hence the only source of dependence between the observed outcomes is the assignment $\mathbf{A}_g$. In the setting with growing group sizes, this assumption can be relaxed at the expense of introducing more restrictions on the within-group network structure (for example, assuming the rate at which links are formed is slow enough). On the other hand, this condition is easily relaxed to arbitrary dependence structures when the group size is fixed. Together, conditions (ii) and (iii) imply that observed outcomes are independent conditional on the assignments. Importantly, note that this condition does not restrict in any way the correlation between treatment assignments $\mathbf{A}_{ig}$ and $\mathbf{A}_{jg}$. In fact, it can be seen that effective treatment assignments are always correlated by construction, since $\mathbf{A}_{ig}$ depends on D_{jg} and vice versa. Finally, part (iv) requires that the total number of parameters not grow faster than the effective sample size, that is, the expected sample size in the smallest cell.

Given a sample of G groups with $n + 1$ units each, let $\mathbb{1}_{ig}(\mathbf{a}) = \mathbb{1}(\mathbf{A}_{ig} = \mathbf{a})$, $N_g(\mathbf{a}) = \sum_{i=1}^{n+1} \mathbb{1}_{ig}(\mathbf{a})$ and $N(\mathbf{a}) = \sum_{g=1}^{G} N_g(\mathbf{a})$, so that $N_g(\mathbf{a})$ is the total number of observations receiving effective assignment $\mathbf{a}$ in group g and $N(\mathbf{a})$ is the total number of observations receiving effective assignment $\mathbf{a}$ in the sample. The estimator for $\mu(\mathbf{a})$ is defined as:

$$
\hat{\mu}(\mathbf{a}) = \begin{cases} \frac{\sum_{g=1}^{G} \sum_{i=1}^{n+1} Y_{ig} \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} & \text{if } N(\mathbf{a}) > 0 \\ \text{undefined} & \text{if } N(\mathbf{a}) = 0 \end{cases}
$$

Thus, the estimator for $\mu(\mathbf{a})$ is simply the sample average of the outcome for observations receiving assignment $\mathbf{a}$, whenever there is at least one observation receiving this assignment.

The following assumption imposes some regularity conditions that are required for upcoming theorems. Let $\sigma^2(\mathbf{a}) = \mathbb{V}[Y_{ig} | \mathbf{A}_{ig} = \mathbf{a}]$.

Assumption 3 (Moments) *There are constants $\underline{\sigma}$ and b such that:*

$$
(i) \quad \inf_n \min_{\mathbf{a} \in \mathcal{A}_n} \sigma^2(\mathbf{a}) \geq \underline{\sigma}^2 > 0, \qquad (ii) \quad \sup_n \max_{\mathbf{a} \in \mathcal{A}_n} \mathbb{E}[Y_{ig}^6 | \mathbf{A}_{ig} = \mathbf{a}] \leq b < \infty
$$

Then we have the following result.

Lemma 2 (Effective sample size) *Suppose Assumptions 2 and 3 hold, and consider an assignment mechanism $\pi(\cdot)$ such that $\pi(\mathbf{a}) > 0$ for all $\mathbf{a} \in \mathcal{A}_n$. If*

$$
\frac{\log |\mathcal{A}_n|}{G \underline{\pi}_n} \to 0 \tag{5}
$$

16

then for any $c \in \mathbb{R}$

$$\mathbb{P}\left[\min_{\mathbf{a} \in \mathcal{A}_n} N(\mathbf{a}) > c\right] \to 1.$$

Lemma 2 says that, under condition (5), the number of observations in the smallest cell will go to infinity, which implies that all the estimators are well defined asymptotically. Hence, condition (5) formalizes the meaning of "large sample" in this context, and states that the number of groups has to be large relative to the total number of parameters and the probability of the least likely assignment. This expression can be interpreted as an invertibility condition for the design matrix of a linear regression model, in the specific case in which the regressors are mutually exclusive indicator variables. This requirement can be seen as a low-level condition that justifies the assumption of invertibility of the design matrix (see e.g. Assumption 2 in Cattaneo, Jansson, and Newey, 2018). When this condition does not hold, small cells may result in estimators with poor finite sample behavior and whose asymptotic distribution, if it exists, may be non-Gaussian. See Ma and Wang (2020) for an example in the context of inverse-probability weighting estimators, which includes randomized experiments as a special case.

Next, let

$$\hat{\sigma}^2(\mathbf{a}) = \frac{\sum_{g=1}^{G} \sum_{i=1}^{n+1} (Y_{ig} - \hat{\mu}(\mathbf{a}))^2 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} \mathbb{1}(N(\mathbf{a}) > 0)$$

be the standard error estimators. Then we have the following result.

Theorem 4 (Consistency and asymptotic normality) *Suppose Assumptions 1, 2 and 3 hold. Under the conditions for Lemma 2,*

$$\max_{\mathbf{a} \in \mathcal{A}_n} |\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})| = O_{\mathbb{P}}\left(\sqrt{\frac{\log |\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),$$

$$\max_{\mathbf{a} \in \mathcal{A}_n} |\hat{\sigma}^2(\mathbf{a}) - \sigma^2(\mathbf{a})| = O_{\mathbb{P}}\left(\sqrt{\frac{\log |\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}\right),$$

(6)

and

$$\max_{\mathbf{a} \in \mathcal{A}_n} \sup_{x \in \mathbb{R}} \left|\mathbb{P}\left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x\right] - \Phi(x)\right| = O\left(\frac{1}{\sqrt{G(n+1)\underline{\pi}_n}}\right)$$

(7)

where $\Phi(x)$ is the cdf of a standard Gaussian random variable.

Equation (6) shows that both the average outcome and standard error estimators converge in probability to their true values, uniformly over treatment assignments, at the rate $\sqrt{\log |\mathcal{A}_n|/(G(n+1)\underline{\pi}_n)}$. The denominator in this rate can be seen as the minimum expected cell size, whereas the numerator is a penalty for having an increasing number of parameters. Equation (7) bounds the difference between the distributions of the standardized average outcomes estimators and the standard normal distribution, uniformly over the treatment

assignments. Under condition (5), $G(n+1)\underline{\pi}_n \to \infty$, which gives asymptotic normality. Furthermore, this bound also reveals the rate at which the distribution of the standardized estimator approaches the standard normal, namely, $\sqrt{G(n+1)\underline{\pi}_n}$.

Another implication of Theorem 4 is that the estimators $\hat{\mu}(\mathbf{a})$ have different convergence rates. More precisely, Lemma A1 in the supplemental appendix and Markov's inequality imply that, for each $\mathbf{a}$, $|\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})| = O_{\mathbb{P}}\left((G(n+1)\pi(\mathbf{a}))^{-1/2}\right)$ and thus the average outcomes for assignments whose probability decreases faster are estimated at a slower rate.

Remark 5 (Inference with many small groups) When the number of units per group is small compared to the total sample size, the effect of group size disappears asymptotically and inference can be based on a large G small n approximation. In this context, n, $|\mathcal{A}_n|$ and $\underline{\pi}_n$ are fixed so condition (5) holds automatically as long as the number of groups goes to infinity. Consistency and asymptotic normality of the estimators can be achieved under the usual regularity conditions as $G \to \infty$, and the variance estimator can easily account for both heteroskedasticity and intragroup correlation using standard techniques. The empirical application in Section 6 fits into this scenario. $\square$

Remark 6 (Inference under misspecified treatment rules) Theorem 4 relies on the true treatment rule $h_0(\cdot)$ begin coarser than $h(\cdot)$, which rules out misspecification of $h(\cdot)$. As discussed in Section 3, in some cases the researcher may want to posit a coarse treatment rule $h(\cdot)$ to use as a summary measure of direct and spillover effects. These cases are not considered in Theorem 4. In such cases, however, as long as $h(\cdot)$ is chosen such that $|\mathcal{A}_n|$ does not increase with the sample size (and therefore the probabilities $\pi(\mathbf{a})$ are fixed), the problem reduces to estimating a finite vector of means. Thus, consistency and asymptotic normality for the vector of parameters of interest follows from existing methods as $G \to \infty$, $n \to \infty$ (or n fixed) and allowing for an arbitrary dependence structure within groups (see e.g. Hansen and Lee, 2019). This applies to cases such as pooled parameters (Remark 2) and partial population experiments (Remark 3). $\square$

Remark 7 (Connection to multi-valued treatments) This asymptotic framework can also be applied to multi-valued treatments setting (Imbens, 2000) where $\mathbf{A}_{ig}$ corresponds to the treatment of unit i in group g. Estimation and inference for multi-valued treatments taking on a finite number of values have been extensively analyzed in a variety of contexts (see e.g. Cattaneo, 2010; Farrell, 2015; Ao, Calonico, and Lee, 2021, and references therein). The results in this section complement this literature in two ways. First, I consider double-array asymptotics in which the number of treatment values is allowed to grow with the sample size. Second, the results account for the dependence between treatment assignments $\mathbf{A}_{ig}$ and $\mathbf{A}_{jg}$ of units in the same group, a feature that is specific of the spillovers setting. $\square$

Remark 8 (Other estimands and uniform inference) Using standard empirical process theory, the estimation and inference results in this section can be extended to hold uniformly on $m(\cdot) \in \mathcal{M}$ for appropriate classes $\mathcal{M}$. See van der Vaart and Wellner (1996) for

a classical reference. For example, letting $\mathcal{M} = \{\mathbb{1}(\cdot \leq y) : y \in \mathbb{R}\}$ leads to the distribution of potential outcomes as discussed in Lemma 1, which can be used to conduct uniform inference on quantile treatment effects and related estimands. $\square$

4.1 Bootstrap approximation

An alternative approach to perform inference in this setting is the bootstrap. Since the challenge for inference is that cells can have too few observations for the Gaussian distribution to provide a good approximation, the wild bootstrap (Shao and Tu, 1995) can offer a more accurate approximation when groups are relatively large. This type of bootstrap can be performed by defining weights $w_{ig} \in \{-1, 1\}$ with probability $1/2$ independently of the sample. The bootstrap estimator for $\mu(\mathbf{a})$ is given by:

$$\hat{\mu}^*(\mathbf{a}) = \frac{\sum_{g=1}^{G} \sum_{i=1}^{n+1} Y_{ig}^* \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}$$

whenever the denominator is non-zero, where

$$Y_{ig}^* \mathbb{1}_{ig}(\mathbf{a}) = (\bar{Y}(\mathbf{a}) + (Y_{ig} - \bar{Y}(\mathbf{a}))w_{ig})\mathbb{1}_{ig}(\mathbf{a}) = (\bar{Y}(\mathbf{a}) + \hat{\varepsilon}_{ig} w_{ig})\mathbb{1}_{ig}(\mathbf{a})$$

In what follows, $\mathbb{P}^*[\cdot]$ denotes a probability calculated over the distribution of w_{ig}, conditional on the sample, and $\mathbb{E}^*[\cdot]$ and $\mathbb{V}^*[\cdot]$ the expectation and variance calculated over $\mathbb{P}^*[\cdot]$. The validity of the wild bootstrap is established in the following theorem.

Theorem 5 (Wild bootstrap) *Under Assumptions 1, 2 and 3,*

$$\max_{\mathbf{a} \in \mathcal{A}_n} \sup_{x \in \mathbb{R}} \left| \mathbb{P}^* \left[\frac{\hat{\mu}^*(\mathbf{a}) - \hat{\mu}(\mathbf{a})}{\sqrt{\mathbb{V}^*[\hat{\mu}^*(\mathbf{a})]}} \leq x \right] - \mathbb{P} \left[\frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \right] \right| \to_{\mathbb{P}} 0.$$

This theorem shows that the wild bootstrap can be used to approximate the distribution of the estimator as an alternative to the standard normal, which may not be accurate when cells have few observations. The performance of the wild bootstrap is illustrated in Section 5 using simulation data.

5 Simulations

This section illustrates the above results in a simulation setting. The outcome is binary and exchangeable, and generated by the following DGP:

$$\mathbb{P}[Y_{ig}(d, s) = 1] = \mu(d, s) = 0.75 + 0.13 \times d + 0.12 \times (1 - d)\mathbb{1}(s > 0)$$

which corresponds to the case with $\mu(0, 0) = 0.75$, $\tau(0) = 0.13$, $\theta_0(s) = 0.12$ for all s and $\theta_1(s) = 0$ for all s. That is, the spillover effect on an untreated unit is equal to 0.12 whenever

at least one neighbor is treated, and zero for treated units. This DGP is based on the empirical application in Section 6.

The simulations consider two assignment mechanisms. First, *simple random assignment* (SR), where treatment is assigned independently with $\mathbb{P}[D_{ig} = 1] = 0.5$. Second, *two-stage randomization with fixed margins* (2SR-FM), where in the first stage groups are assigned the total number of treated units between $t = 0, 1, 2, \ldots, n + 1$ with probability q_t, and then treated units are chosen randomly within each group. The probabilities q_t are chosen so that sample sizes across all assignments are similar; see Section A3 in the supplemental appendix for further details. Corollaries A1 and A2 in the supplemental appendix show that when peers are exchangeable, under SR, condition (5) from Theorem 4 holds whenever $(n + 1)/\log G \to 0$, whereas under 2SR-FM the condition holds when $\log(n + 1)/\log G \to 0$. Because the second condition is weaker, the normal approximation is expected to perform better for 2SR-FM as groups get larger.

The parameter of interest will be $\theta_0(n) = \mathbb{E}[Y_{ig}(0, n)] - \mathbb{E}[Y_{ig}(0, 0)]$, which is the average spillover effect for an untreated unit with all peers treated. In this simulation, $\theta_0(n) = 0.12$. This parameters can be seen as a "worst-case scenario" given that the probability of the assignment $(D_{ig}, S_{ig}) = (0, n)$ is one of the smallest (in fact, the smallest under 2SR-FM). The spillover effect estimator is the difference in cell means:

$$\hat{\theta}_0(n) = \frac{\sum_{g=1}^{n+1} \sum_{i=1}^{n+1} Y_{ig} \mathbb{1}_{ig}(0, n)}{N(0, n)} - \frac{\sum_{g=1}^{n+1} \sum_{i=1}^{n+1} Y_{ig} \mathbb{1}_{ig}(0, 0)}{N(0, 0)}$$

whenever $N(0, n) > 1$ and $N(0, 0) > 1$, so that both the estimator and its standard error can be calculated. When at least one of the cells has one or zero observations, the estimator is undefined.

Table 1 presents the results for a sample with 300 groups, for group sizes, $n + 1 \in \{3, 4, \ldots, 8\}$. The upper panel shows the results under SR while the lower panel corresponds to the 2SR-FM assignment. In each panel, the first row gives the value of the condition to achieve consistency and asymptotic normality; intuitively, the closer this value is to zero, the better the Gaussian approximation should be. The second and third rows show the bias and the variance of $\hat{\theta}_0(n)$, calculated over the values of the simulated estimates conditional on the estimate being well defined (i.e. when the cells have enough observations to calculate the estimator and its variance). Rows four to seven show the coverage rate and average length of a 95% confidence intervals based on the Gaussian approximation and the wild bootstrap. The eighth row gives the proportion of the simulations in which the estimator or its standard error could not be calculated due to insufficient number of observations. Finally the last two rows show the average sample size in the two assignment cells of interest. Coverage rates are also shown in Figure 1.

The simulations reveal that under simple randomization, the estimator performs well for $n \leq 4$, with no bias and coverage rates very close to 95% for both the asymptotic normal and the wild bootstrap. When $n > 4$, however, coverage rates decrease in both cases, more

Table 1: Simulation results, $G = 300$

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
Simple randomization						
$(n+1)/\log(G)$	0.5260	0.7013	0.8766	1.0519	1.2273	1.4026
Bias	-0.0022	-0.0003	0.0003	-0.0009	-0.0033	0.0011
Variance	0.0027	0.0042	0.0070	0.0128	0.0214	0.0307
95% CI coverage - normal	0.9488	0.9460	0.9406	0.9344	0.9090	0.8872
95% CI length - normal	0.2041	0.2522	0.3238	0.4293	0.5516	0.6571
95% CI coverage - bootstrap	0.9504	0.9474	0.9430	0.9391	0.9221	0.9275
95% CI length - bootstrap	0.2044	0.2536	0.3291	0.4443	0.5891	0.7342
Prop. empty cells	0.0000	0.0000	0.0000	0.0122	0.0946	0.3154
$\mathbb{E}[N(0,0)]$	112	75	47	28	16	9
$\mathbb{E}[N(0,n)]$	112	75	47	28	16	9
Two-stage randomization						
$\log(n+1)/\log(G)$	0.1926	0.2430	0.2822	0.3141	0.3412	0.3646
Bias	0.0000	0.0003	0.0001	0.0000	-0.0005	0.0010
Variance	0.0027	0.0031	0.0035	0.0039	0.0042	0.0046
95% CI coverage - normal	0.9518	0.9456	0.9472	0.9510	0.9468	0.9412
95% CI length - normal	0.2036	0.2166	0.2319	0.2422	0.2531	0.2616
95% CI coverage - bootstrap	0.9524	0.9470	0.9490	0.9512	0.9452	0.9424
95% CI length - bootstrap	0.2037	0.2172	0.2327	0.2432	0.2548	0.2636
Prop. empty cells	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
$\mathbb{E}[N(0,0)]$	113	100	89	82	76	72
$\mathbb{E}[N(0,n)]$	112	100	88	82	76	72

Notes: simulation results for $G = 300$ groups. The second and third rows in each panel show the bias and variance of $\hat{\theta}_n(0)$. The fourth to seventh rows show the coverage rate and average length of a normal-based and wild-bootstrap-based 95% confidence intervals, respectively. The eighth row shows the proportion of simulations in which $\hat{\theta}_n(0)$ is undefined due to the small number of observations in the corresponding cell. The ninth and tenth rows show the average sample size in the two assignment cells of interest. Results from 5,000 simulations with 1,000 bootstrap repetitions.

Table 2: Simulation results, $G = 600$

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
Simple randomization						
$(n+1)/\log(G)$	0.4690	0.6253	0.7816	0.9379	1.0943	1.2506
Bias	-0.0004	0.0001	0.0002	0.0003	0.0010	0.0012
Variance	0.0013	0.0020	0.0033	0.0058	0.0110	0.0186
95% CI coverage - normal	0.9522	0.9474	0.9514	0.9492	0.9381	0.9237
95% CI length - normal	0.1437	0.1766	0.2250	0.2958	0.3992	0.5171
95% CI coverage - bootstrap	0.9496	0.9484	0.9488	0.9506	0.9423	0.9388
95% CI length - bootstrap	0.1433	0.1767	0.2258	0.2989	0.4121	0.5456
Prop. empty cells	0.0000	0.0000	0.0000	0.0000	0.0088	0.0958
$\mathbb{E}[N(0,0)]$	225	150	94	56	33	19
$\mathbb{E}[N(0,n)]$	225	150	94	56	33	19
Two-stage randomization						
$\log(n+1)/\log(G)$	0.1717	0.2167	0.2516	0.2801	0.3042	0.3251
Bias	-0.0001	-0.0003	0.0009	-0.0003	0.0006	-0.0003
Variance	0.0013	0.0015	0.0017	0.0019	0.0020	0.0022
95% CI coverage - normal	0.9506	0.9504	0.9502	0.9476	0.9442	0.9498
95% CI length - normal	0.1437	0.1526	0.1631	0.1695	0.1768	0.1817
95% CI coverage - bootstrap	0.9498	0.9488	0.9498	0.9460	0.9448	0.9490
95% CI length - bootstrap	0.1435	0.1523	0.1629	0.1694	0.1768	0.1818
Prop. empty cells	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathbb{E}[N(0,0)]$	225	200	177	164	152	144
$\mathbb{E}[N(0,n)]$	225	200	177	164	152	144

Notes: simulation results for $G = 600$ groups. The second and third rows in each panel show the bias and variance of $\hat{\theta}_n(0)$. The fourth to seventh rows show the coverage rate and average length of a normal-based and wild-bootstrap-based 95% confidence intervals, respectively. The eighth row shows the proportion of simulations in which $\hat{\theta}_n(0)$ is undefined due to the small number of observations in the corresponding cell. The ninth and tenth rows show the average sample size in the two assignment cells of interest. Results from 5,000 simulations with 1,000 bootstrap repetitions.
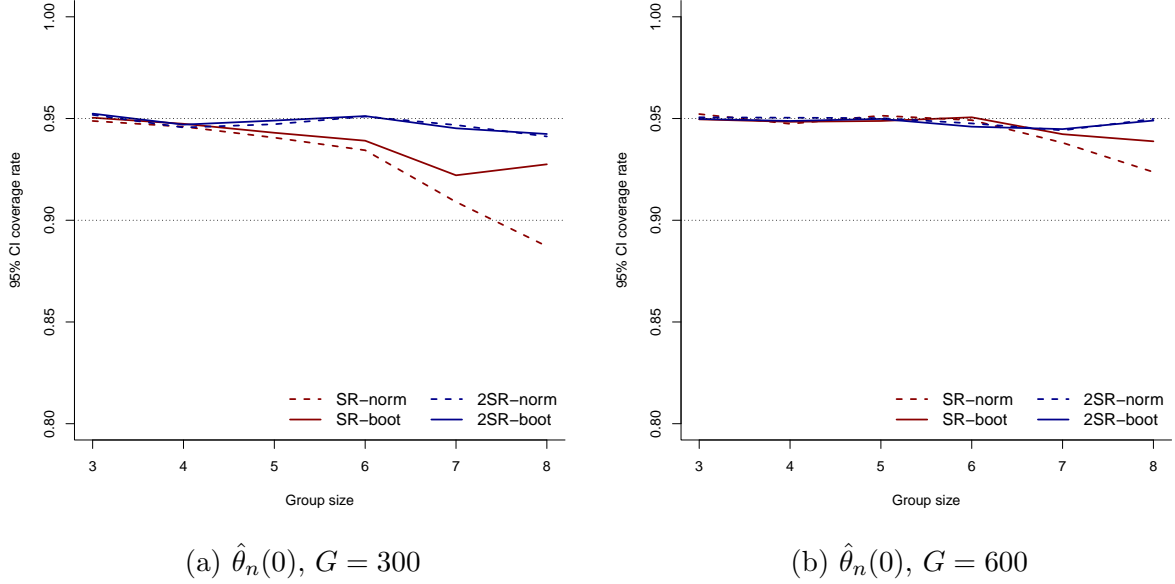
Figure 1: Coverage rate of the 95% confidence interval.

(a) $\hat{\theta}_n(0)$, $G = 300$

(b) $\hat{\theta}_n(0)$, $G = 600$

Notes: the dashed lines show the coverage rate of the 95% confidence interval for $\theta_n(0)$ based on the normal approximation under simple random assignment (red line) and two-stage randomization (blue line) for a sample with 300 (left) and 600 (right) groups. The solid lines show the coverage rates for the confidence interval constructed using wild bootstrap.

rapidly so for the normal confidence interval whose coverage rate drops to 88% for $n = 7$. While the coverage of the bootstrap confidence interval also decreases, it remains closer to 95% compared to the asymptotic normal approximation. The table also shows that under this assignment mechanism, the corresponding sample sizes decrease very rapidly in the relevant assignment cells. When $n = 7$, each mean is calculated using 9 observations on average, and is undefined in about 32% of the simulations.

On the other hand, under two-stage randomization, both the asymptotic normal and the wild bootstrap confidence intervals perform equally well and coverage remains very close to 95%. As shown in the last two rows, two-stage randomization ensures much larger sample sizes in the assignment cells of interest compared to simple randomization.

Table 2 shows the same results for a sample with 600 groups. As expected, the estimator and confidence intervals show better performance compare to the case with $G = 300$.

6 Empirical illustration

In this section I reanalyze the data from Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011). The authors conducted a pilot experiment designed to evaluate the effect of a conditional cash transfer program, *Subsidios Condicionados a la Asistencia Escolar*, in Bogotá, Colombia. The program aimed at increasing student retention and reducing drop-out and

Table 3: Distribution of household size and number of treated per household

	HH size	Treated per HH
0	0	2,355
1	5,205	3,782
2	1,410	607
3	168	52
4	15	3
5	1	0
Total	6,799	6,799

Notes: the first column indicates the frequencies of household size (i.e. the number of eligible children per household) and the second column indicates the frequencies of the number of treated children per household in the sample collected by Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011).

child labor. Eligible registrants ranging from grade 6-11 were randomly assigned to treatment and control.[2] The assignment was performed at the student level. In addition to administrative and enrollment data, the authors collected baseline and follow-up data from students in the largest 68 of the 251 schools. This survey contains attendance data and was conducted in the household. As shown in the first column of Table 3, 1,594 households have more than one registered child (rows labeled 2 to 5), and since the treatment was assigned at the child level, this gives variation in the number of treated children per household. This can be seen in the second column of Table 3, which shows that the number of treated children varies from 0 to 5.

I analyze direct and spillover effects restricting the sample to households with three registered siblings, which gives a total of 168 households and 504 observations. The outcome of interest is school attendance. Because groups are very small in this case, inference can be conducted using standard methods (see Remark 5).

I start by estimating the average direct and spillover effects exploiting variation in the number of treated siblings using the following regression:

$$\mathbb{E}[Y_{ig}|D_{ig},S_{ig}] = \alpha + \tau D_{ig} + \sum_{s=1}^{n_g} \theta_0(s)\mathbb{1}(S_{ig}=s)(1-D_{ig}) + \sum_{s=1}^{n_g} \theta_1(s)\mathbb{1}(S_{ig}=s)D_{ig} \quad (8)$$

Because this regression is saturated, it follows that:

$$\tau = \mathbb{E}[Y_{ig}|D_{ig}=1,S_{ig}=0] - \mathbb{E}[Y_{ig}|D_{ig}=0,S_{ig}=0]$$

and

$$\theta_d(s) = \mathbb{E}[Y_{ig}|D_{ig}=d,S_{ig}=s] - \mathbb{E}[Y_{ig}|D_{ig}=d,S_{ig}=s].$$

Lemma 1 provides two alternative ways to interpret these estimands. If siblings are exchange-

---

[2]The experiment had two different treatments that varied the timing of the payments, but, following the authors, I pool the two treatment arms to increase the sample size. See Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011) for details.

able, so that average potential outcomes take the form $\mathbb{E}[Y_{ig}(d, s)]$, then $\tau = \mathbb{E}[Y_{ig}(1,0) - Y_{ig}(0,0)]$ is the average direct effect of the treatment on a child with no treated siblings, whereas $\theta_d(s) = \mathbb{E}[Y_{ig}(d, s) - Y_{ig}(d, 0)]$ is the average spillover effect of having s treated siblings under own treatment status d. In this application, exchangeability may be a reasonable assumption if parents make schooling decisions based on how many of their children are treated (for example, to determine whether the cash transfer covers the direct and opportunity costs of sending their children to school), regardless of which of their children are treated.

On the other hand, if exchangeability does not hold, Lemma 1 shows that the parameters from Equation (8) combine weighted averages of average potential outcomes. For example, if siblings have a certain (possibly unknown) ordering so that the true average potential outcomes take the form $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)]$, then:

$$\tau = \mathbb{E}[Y_{ig}(1, \mathbf{0}_g) - Y_{ig}(0, \mathbf{0}_g)]$$

and

$$\theta_d(s) = \sum_{\mathbf{d}_g : \mathbf{1}_g' \mathbf{d}_g = s} \mathbb{E}[Y_{ig}(d, \mathbf{d}_g) - Y_{ig}(d, \mathbf{0}_g)]\mathbb{P}[\mathbf{D}_{(i)g} = \mathbf{d}_g | D_{ig} = d, S_{ig} = s]$$

which averages the spillover effects over the values of $\mathbf{D}_{(i)g}$ that are consistent with $S_{ig} = s$. Thus, Equation (8) provides a way to summarize the direct and spillover effects of the program even when the true structure of the average potential outcomes is unknown. In addition, Tables 5 and 6 provide additional results that explore this issue further and provide a way to test exchangeability over different dimensions.

The estimates from Equation (8) are shown in the third panel, "**Full**", of Table 4. These estimates suggest a positive direct effect of the treatment of 16.4 percentage points, significant at the 5 percent level, with almost equally large spillover effects on the untreated units. More precisely, the estimated effect on an untreated kid of having one treated sibling is 14.6 percentage points, while the effect of having two treated siblings is 14 percentage points. The fact that we cannot reject the hypothesis that $\theta_0(1) = \theta_0(2)$ suggests some form of crowding-out: given that one sibling is treated, treating one more sibling does not affect attendance. These facts are consistent with the idea that, for example, the conditional cash transfer alleviates some financial constraint that was preventing the parents from sending their children to school regularly, or with the program increasing awareness on the importance of school attendance, since in these cases the effect occurs as soon as at least one kid in the household is treated, and does not amplify with more treated children.

On the other hand, spillover effects on treated children are much smaller in magnitude and negative. The fact that these estimates are negative does not mean that the program hurts treated children, but that treating more siblings reduces the benefits of the program. For example, the effect of being treated with two treated siblings, compared to nobody treated, can be estimated by $\hat{\tau} + \hat{\theta}_1(2) \approx 0.113$. Thus, a treated kid with two treated siblings

25

increases her attendance in 11 percentage points starting from a baseline in which nobody in the household is treated.

In all, the estimates suggest large and positive direct and spillover effects on the untreated, with some evidence of crowding-out between treated siblings.[3]

6.1 Difference in means

Using the results from the nonparametric specification as a benchmark, I now estimate the effects of the program using the difference in means analyzed in Section 3.1. The left panel, "**Diff. Means**", of Table 4 shows the difference in means, calculated as the OLS estimator for β_{D} in Equation (1). The results show that the difference in means is practically zero and not significant. Hence, by ignoring the presence of spillover effects, a researcher estimating the effect of the program in this way would conclude that the treatment has no effect.

This finding is due to the fact that the difference in means combines all the effects in the third panel into a single number, as shown in Theorem 1. From Table 4, the estimated spillover effects in this case are larger under control that under treatment, and have different signs, so $\hat{\theta}_1(s) - \hat{\theta}_0(s) < 0$. Therefore, the spillover effects push the difference in means towards zero in this case.

6.2 Reduced-form linear-in-means regression

Next, I estimate the effects using the RF-LIM regression analyzed in Section 3.2. The estimates from Equation (2) are given in the first column of the middle panel, "**Linear-in-Means**", in Table 4. The estimates reveal very small and not significant direct and spillover effects, substantially different from the results using Equation (8).

Theorem 2 and Corollary 1 show that a RF-LIM regression implicitly imposes linearity of spillover effects and may suffer from misspecification when spillover effects are nonlinear. The estimates from the full nonparametric specification shows that spillover effects are highly nonlinear in this case, which explains why the RF-LIM regression fails to recover these effects.

The second column in the RF-LIM panel presents the estimates from the interacted RF-LIM regression shown in Equation (3). The results reveal that separately estimating the spillover effects for treated and controls mitigates the misspecification in this case, and the estimates are closer to the ones from the nonparametric specification, although the fact that 0.169 is not a weighted average of 0.146 and 0.14 suggests that some extrapolation bias remains due to the nonlinearity of spillover effects.

---

[3]These empirical findings differ from those in Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011), who find evidence of negative spillover effects. The latter results are calculated over a different sample, since the authors focus on households with two registered children, whereas I consider households with three registered children. Differences in estimated direct and spillover effects may be due to heterogeneous effects across household sizes. For instance, in this sample, households with more registered children have lower income on average, so they may benefit differently from the program.

Table 4: Estimation results

	Diff. Means		Linear-in-Means				Full		Pooled	
	coef	s.e.	coef	s.e.	coef	s.e.	coef	s.e.	coef	s.e.
D_{ig}	0.006	0.016	0.007	0.016	0.102**	0.042	0.164**	0.066	0.165**	0.065
$\bar{D}_g^{(i)}$			0.027	0.034						
$\bar{D}_g^{(i)}(1 - D_{ig})$					0.169**	0.068				
$\bar{D}_g^{(i)} D_{ig}$					-0.064	0.039				
$\mathbb{1}(S_{ig} = 1)(1 - D_{ig})$							0.146**	0.066		
$\mathbb{1}(S_{ig} = 2)(1 - D_{ig})$							0.14**	0.056		
$\mathbb{1}(S_{ig} = 1)D_{ig}$							-0.041*	0.023		
$\mathbb{1}(S_{ig} = 2)D_{ig}$							-0.051**	0.025		
$\mathbb{1}(S_{ig} > 0)(1 - D_{ig})$									0.144**	0.06
$\mathbb{1}(S_{ig} > 0)D_{ig}$									-0.045**	0.02
Constant	0.822***	0.013	0.811***	0.024	0.756***	0.037	0.706***	0.057	0.705***	0.057
Observations		504		504		504		504		504

Notes: S.e. clustered at the household level. Regressions include school FE. ***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

6.3 Pooled effects

I now illustrate how to estimate pooled effects by averaging over the possible number of treated siblings (2 and 3 in this case). For this, I estimate the following regression:

$$Y_{ig} = \alpha_p + \tau D_{ig} + \gamma_p^0 \mathbb{1}(S_{ig} > 0)(1 - D_{ig}) + \gamma_p^1 \mathbb{1}(S_{ig} > 0)D_{ig} + \nu_{ig}$$

where

$$\tau = \mathbb{E}[Y_{ig}|D_{ig} = 1, S_{ig} = 0] - \mathbb{E}[Y_{ig}|D_{ig} = 0, S_{ig} = 0]$$

and

$$\gamma_p^d = \sum_{s=1}^{2} \theta_d(s)\mathbb{P}[S_{ig} = s|D_{ig} = d, S_{ig} > 0]$$

where $\theta_d(s)$ is defined in Equation (8) (see also Remark 2). From Table 4 we can see that the estimated pooled spillover effects are 0.144 for controls and -0.045 for treated, which as expected lie between the effects found with the saturated regression. These results illustrate how this type of pooling can provide a useful summary of spillover effects, which may be a feasible alternative when the total number of spillover effects is too large or cell sizes are small to estimate them separately.

6.4 Non-exchangeable peers

Next, I illustrate how to relax the exchangeability assumption in two ways. First, I define an ordering between siblings by looking at differences (in absolute value) in ages, defining sibling 1 as the sibling closest in age and sibling 2 as the sibling farthest in age. Then, estimation is conducted by simply adding indicator variables for the possible different assignments. Table 5 shows the estimates from this specification. The estimates reveal similar results to Table 4, with a direct effect of 0.165, spillover effects on the untreated ranging from 0.133 to 0.14 and spillover effects on the treated ranging from -0.039 to -0.051.

In fact, exchangeability can be tested by assessing whether the spillover effects of siblings 1 and 2 are the same, which in this case amounts to testing equality of coefficients between rows 2 and 3, and between rows 5 and 6. The test statistic and corresponding p-value from this test are given in the last two rows of the table, where it is clear that exchangeability cannot be rejected in this case, although this could be due to low statistical power given the relatively small sample size.

Finally, I consider the case in which the effect of treated siblings depends on whether siblings are male or female, allowing for non-exchangeable siblings based on gender. The results are shown in Table 6. In this table, S_{ig}^m denotes the number of male treated siblings and S_{ig}^f denotes the number of female treated siblings. The results are qualitatively similar, with some suggestive evidence of slightly larger spillover effects from female siblings. The hypothesis that the coefficients are the same cannot be rejected, although again the sample may be too small to draw precise conclusions about sibling exchangeability.

Table 5: Estimation results by age

	coef	s.e.
D_{ig}	0.165**	0.066
$(1 - D_{ig})D_{i1g}(1 - D_{i2g})$	0.132**	0.067
$(1 - D_{ig})(1 - D_{i1g})D_{i2g})$	0.163**	0.069
$(1 - D_{ig})D_{i1g}D_{i2g}$	0.141**	0.056
$D_{ig}D_{i1g}(1 - D_{i2g})$	-0.041	0.027
$D_{ig}(1 - D_{i1g})D_{i2g}$	-0.041	0.026
$D_{ig}D_{i1g}D_{i2g}$	-0.051**	0.025
Constant	0.706***	0.057
Observations		504
Chi-squared test	0.482	
p-value	0.619	

Notes: Cluster-robust s.e. Regressions include school FE. ***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

7 Discussion

The findings in this paper offer several takeaways for analyzing spillover effects in randomized experiments. First, commonly-analyzed estimands such as the difference in means and RF-LIM coefficients implicitly impose strong assumptions on the structure of spillover effects, and are therefore not recommended as they generally do not have a causal interpretation. On the other hand, the full vector of spillover effects is identifiable whenever the design generates enough variation in the number of treated units in each group and the researcher assumes a treatment rule that is flexible enough.

Second, while nonparametric estimation of all direct and spillover effects can give a complete picture of the effects of the treatment, it can be difficult to implement in practice when groups are large. As a guideline to determine in which cases spillover effects can be estimated nonparametrically, Theorem 4 formalizes the notion of a "large enough sample" in this context, and provides a way to assess the performance of the different types of treatment effect estimators depending on the number of groups, number of parameters of interest and treatment assignment mechanism. As an alternative, pooled estimands can recover weighted averages of spillover effects with known weights for which inference can be conducted under mild conditions.

The supplemental appendix discusses several important issues that can be further developed in future research. Sections A4 and A5 discuss extensions to unequal group sizes and the inclusion of covariates. The results in Section 4 and the simulations in Section 5 highlight the fact that the rate of convergence of the proposed estimators depend on the design. This suggests that these results can be used to rank treatment assignment mechanisms, and this has implications for experimental design, as discussed in Section A2.

The analysis in this paper leaves several open questions to be explored. One example

Table 6: Estimation results by gender

	coef	s.e.
D_{ig}	0.124***	0.045
$\mathbb{1}(S_{ig}^m = 1)(1 - D_{ig})$	0.035	0.036
$\mathbb{1}(S_{ig}^m = 2)(1 - D_{ig})$	0.105**	0.045
$\mathbb{1}(S_{ig}^m = 1)D_{ig}$	-0.013	0.024
$\mathbb{1}(S_{ig}^m = 2)D_{ig}$	-0.032	0.025
$\mathbb{1}(S_{ig}^f = 1)(1 - D_{ig})$	0.097***	0.035
$\mathbb{1}(S_{ig}^f = 2)(1 - D_{ig})$	0.101**	0.042
$\mathbb{1}(S_{ig}^f = 1)D_{ig}$	-0.067***	0.025
$\mathbb{1}(S_{ig}^f = 2)D_{ig}$	0.000	0.026
Constant	0.745***	0.04
Observations		504
Chi-squared test	1.578	
p-value	0.182	

Notes: Cluster-robust s.e. Regressions include school FE. ***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

is allowing for endogenous group formation. The identification results in this paper follow through when groups are endogenously formed, as long as they are formed before the treatment is assigned and their structure is not changed by the treatment (inference may require further assumptions to account for possible overlap between groups). On the other hand, when the structure of the group is affected by the treatment, the treatment can affect outcomes through direct effects, through spillover effects given the network, and through changing the network structure. In such cases, while it is possible to identify the "overall" effect (Kline and Tamer, 2019), random assignment of the treatment is generally not enough to separately point identify these different effects, and further assumptions are needed. On the other hand, the findings in this paper can be generalized to settings where the researcher does not have precise control on treatment receipt. In this direction, Vazquez-Bare (2021) analyzes instrumental variable methods that can be applied to RCTs with imperfect compliance, where treatment receipt is endogenous, or more generally in quasi-experimental settings.

References

ABADIE, A., AND M. D. CATTANEO (2018): "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

AO, W., S. CALONICO, AND Y.-Y. LEE (2021): "Multivalued Treatments and Decomposition Analysis: An Application to the WIA Program," *Journal of Business & Economic Statistics*, 39(1), 358–371.

ATHEY, S., D. ECKLES, AND G. W. IMBENS (2018): "Exact P-values for Network Interference," *Journal of the American Statistical Association*, 113(521), 230–240.

ATHEY, S., AND G. IMBENS (2017): "The Econometrics of Randomized Experiments," in *Handbook of Field Experiments*, ed. by A. V. Banerjee, and E. Duflo, vol. 1 of *Handbook of Economic Field Experiments*, pp. 73 – 140. North-Holland.

BAIRD, S., A. BOHREN, C. MCINTOSH, AND B. ÖZLER (2018): "Optimal Design of Experiments in the Presence of Interference," *The Review of Economics and Statistics*, 100(5), 844–860.

BARRERA-OSORIO, F., M. BERTRAND, L. L. LINDEN, AND F. PEREZ-CALLE (2011): "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia," *American Economic Journal: Applied Economics*, 3(2), 167–195.

BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND R. JAYARAMAN (2015): "Linear Social Interactions Models," *Journal of Political Economy*, 123(2), 444–496.

BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150(1), 41 – 55.

CATTANEO, M. D. (2010): "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138 – 154.

CATTANEO, M. D., M. JANSSON, AND W. K. NEWEY (2018): "Inference in Linear Regression Models with Many Covariates and Heteroskedasticity," *Journal of the American Statistical Association*, 113(523), 1350–1361.

CRÉPON, B., E. DUFLO, M. GURGAND, R. RATHELOT, AND P. ZAMORA (2013): "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment," *The Quarterly Journal of Economics*, 128(2), 531–580.

DAVEZIES, L., X. D'HAULTFOEUILLE, AND D. FOUGÈRE (2009): "Identification of peer effects using group size variation," *Econometrics Journal*, 12(3), 397–413.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010): "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2(2), 241–75.

Duflo, E., and E. 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.

Farrell, M. H. (2015): "Robust inference on average treatment effects with possibly more covariates than observations," *Journal of Econometrics*, 189(1), 1 – 23.

Goldsmith-Pinkham, P., and G. W. Imbens (2013): "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, 31(3), 253–264.

Halloran, M. E., and M. G. Hudgens (2016): "Dependent Happenings: a Recent Methodological Review," *Current Epidemiology Reports*, 3(4), 297–305.

Hansen, B. E., and S. Lee (2019): "Asymptotic theory for clustered samples," *Journal of Econometrics*, 210(2), 268–290.

Hirano, K., and J. Hahn (2010): "Design of Randomized Experiments to Measure Social Interaction Effects," *Economics Letters*, 106(1), 51–53.

Hudgens, M. G., and M. E. Halloran (2008): "Toward Causal Inference with Interference," *Journal of the American Statistical Association*, 103(482), 832–842.

Imbens, G. (2000): "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87(3), 706–710.

Imbens, G. W., and D. B. Rubin (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kline, B., and E. Tamer (2019): "Econometric Analysis of Models with Social Interactions," in *The Econometric Analysis of Network Data*, ed. by B. S. Graham, and A. De Paula. Elsevier.

Lazzati, N. (2015): "Treatment response with social interactions: Partial identification via monotone comparative statics," *Quantitative Economics*, 6(1), 49–83.

Lee, L.-F. (2007): "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 140(2), 333–374.

Ma, X., and J. Wang (2020): "Robust Inference Using Inverse Probability Weighting," *Journal of the American Statistical Association*, 115(532), 1851–1860.

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

———— (2013a): "Comment," *Journal of Business & Economic Statistics*, 31(3), 273–275.

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

MOFFIT, R. (2001): "Policy Interventions, Low-level Equilibria and Social Interactions," in *Social Dynamics*, ed. by S. N. Durlauf, and P. Young, pp. 45–82. MIT Press.

OGBURN, E. L., AND T. J. VANDERWEELE (2014): "Causal Diagrams for Interference," *Statistical Science*, 29(4), 559–578.

SHAO, J., AND D. TU (1995): *The Jackknife and Bootstrap*. Springer New York, New York, NY.

SOBEL, M. E. (2006): "What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference," *Journal of the American Statistical Association*, 101(476), 1398–1407.

TCHETGEN TCHETGEN, E. J., AND T. J. VANDERWEELE (2012): "On causal inference in the presence of interference," *Statistical Methods in Medical Research*, 21(1), 55–75.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer New York, New York, NY.

VAZQUEZ-BARE, G. (2021): "Causal Spillover Effects Using Instrumental Variables," *arXiv:2003.06023*.