# Identification and Estimation of Spillover Effects in Randomized Experiments: Supplemental Appendix

Gonzalo Vazquez-Bare*

July 22, 2020

**Abstract**

This supplemental appendix provides the proofs of the results in the paper and additional discussions and results not included in the paper to conserve space.

## Contents



*Department of Economics, University of California, Santa Barbara. gvazquez@econ.ucsb.edu

# A1 Review of the literature

For a direct comparison between the literature and my setup, I will focus on the case in which the covariate of interest is a binary treatment $D_{ig}$. I will follow the notation introduced in the paper.

## A1.1 Econometric models, LIM, response functions

The linear-in-means (LIM) model is arguably the most widely used tool to analyze peer effects in most areas of economics. In its standard version, a LIM model is given by the equation:

$$Y_{ig} = \alpha + \beta D_{ig} + \gamma \bar{D}_g^{(i)} + \delta \bar{Y}_g^{(i)} + \varepsilon_{ig} \tag{1}$$

where $\bar{D}_g^{(i)}$ is the leave-one-out sample average of $D$ (in this case, the proportion of ones) excluding $D_{ig}$, and similarly for $\bar{Y}_g^{(i)}$. In this equation, $\beta$ is the direct effect of the treatment, $\gamma$ is the *exogenous effect* and $\delta$ is the *endogenous effect*. A "large group" version of this equation replaces $\bar{D}_g^{(i)}$ and $\bar{Y}_g^{(i)}$ with their within-group population averages $\mathbb{E}_g[D]$ and $\mathbb{E}_g[Y]$ (see e.g. Manski, 1993). A LIM model is often interpreted as the Nash equilibrium of a game in which players maximize a quadratic utility function (Blume, Brock, Durlauf, and Jayaraman, 2015; Kline and Tamer, 2019):

$$\max_{Y_{ig}} U\left(Y_{ig}, \bar{Y}_g^{(i)}\right) = \xi_{ig} Y_{ig} - \frac{(1-\delta)Y_{ig}^2}{2} - \frac{\delta}{2}(Y_{ig} - \bar{Y}_g^{(i)})^2$$

In this equation, the first two terms represent a private component of utility, with marginal private benefit equal to $\xi_{ig}$ and a convex cost, and the last term captures a "social pressure" component (Blume, Brock, Durlauf, and Jayaraman, 2015). The presence of this last term implies that an individual gets higher utility by choosing an action $Y_{ig}$ that is close the the average action in her group. The first-order condition of this maximization problem yields Equation 1 after setting $\xi_{ig} = \alpha + \beta D_{ig} + \gamma \bar{D}_g^{(i)} + \varepsilon_{ig}$.

Manski (1993) pointed out two identification problems associated with model 1. First, the model includes endogenous variables, namely, the outcomes of other units, as regressors (the *reflection problem*). Second, the presence of a group-level fixed effect can generate a correlation between the error term and the regressors (the problem of *correlated effects*). Several approaches have been put forward to ensure identification of $\gamma$ and $\delta$, such as exploiting the variation generated by partially-overlapping groups (Bramoullé, Djebbari, and Fortin, 2009; De Giorgi, Pellizzari, and Redaelli, 2010), using variation in group sizes (Lee, 2007; Davezies, D'Haultfoeuille, and Fougère, 2009) or combining the availability of an exogenous instrument with the panel-like structure of the data (Graham and Hahn, 2005). However, these methods only work in specific contexts and can be very demanding in terms of data requirements. A more straightforward approach taken by the literature is to give up separate identification of $\gamma$ and $\delta$, and use the fact that under appropriate restrictions on the model

parameters, Equation 1 can be solved to obtain a reduced-form equation (corresponding to a Nash equilibrium):

$$Y_{ig} = \lambda + \mu D_{ig} + \theta \bar{D}_g^{(i)} + u_{ig} \tag{2}$$

where the coefficients $(\lambda, \mu, \theta)$ are (nonlinear) functions of the structural parameters $(\alpha, \beta, \gamma, \delta)$. In this case, $\theta$ captures the composite exogenous and endogenous peer effect. Although Equation 2 does not allow separate identification of the exogenous and endogenous effects, Manski (2013a,b) has argued that the reduced form may actually be the object of interest from a policy perspective, since a policy intervention can affect exogenous variables but not outcomes directly.

While Equation 2 circumvents the endogeneity generated by the presence of $\bar{Y}_g^{(i)}$, its parameters remain unidentified when $(D_{ig}, \bar{D}_g^{(i)})$ are correlated with $u_{ig}$. Such correlation can arise, for example, when units in the same group are subject to common shocks. If these common shocks are correlated with the regressors, the reduced-form parameters are not identified. For instance, suppose $D_{ig}$ indicates whether student $i$ in classroom $g$ has ever failed a grade, and $\bar{D}_g^{(i)}$ is the proportion of students excluding $i$ that have failed a grade (repeaters). If classrooms with a higher proportion of repeaters are assigned better teachers, then teacher quality is a group-level shock that is correlated with the regressors, and it is impossible to disentangle the effect of having more repeaters from the effect of having better teachers.

Again, the literature has offered several alternatives to deal with this issue. A credible approach has been to rely on random assignment to eliminate the correlation between the regressors and the error term. There are two main ways in which randomization is conducted in the peer effects literature. The first one is random assignment of group membership. For instance, Sacerdote (2001), Zimmerman (2003) and Stinebrickner and Stinebrickner (2006) exploit random (or random-like) assignment of college roommates, while Lyle (2007) and Carrell, Fullerton, and West (2009) study the case of random assignment into peer groups in West Point and the Air Force Academy, respectively. Graham (2008) takes advantage of the random assignment of students to small and large classrooms in Project STAR to identify peer effects through variance contrasts. However, random assignment of groups breaks apart when individuals refuse to interact with the peers they were assigned to (Carrell, Sacerdote, and West, 2013). The second method is direct random assignment of the treatment. Moffit (2001) argued in favor of *partial-population experiments*, in which the proportion of treated units is randomized in each group. Some examples exploiting random assignment of treatment in linear-in-means models are Lalive and Cattaneo (2009), Bobonis and Finan (2009) and Dieye, Djebbari, and Barrera-Osorio (2014).

Even when randomization is possible, identification of the parameters still relies strongly on the linearity imposed on Equations 1 and 2, and the question remains of whether (i) the linear model is an appropriate representation of the phenomenon under study and (ii) it is possible to achieve identification without imposing a linear structure. Attempts to relax the linearity assumption have been motivated by both theoretical and empirical considerations.

On the one hand, the linear model is generally incorrect when outcomes are binary or discrete. This observation sparked a large literature on binary-choice models with social interactions (see e.g. Brock and Durlauf, 2001). Although this literature removes linearity, it usually does so by replacing it by alternative (and possibly equally strong) parametric or distributional assumptions. On the other hand, the linear model has been criticized on empirical grounds for the unrealistic restrictions that it imposes on the structure of peer effects (Hoxby and Weingarth, 2005; Carrell, Fullerton, and West, 2009; Sacerdote, 2011, 2014).

On the opposite end of the spectrum, Manski (2013b) and Lazzati (2015) study nonparametric partial identification of the response function, that is, the reduced-form relationship between outcomes and treatment assignments, in presence of social interactions. These papers characterize identification regions for the distribution of the potential outcomes under different restrictions on the structural model, the response functions and the structure of social interactions.

In this paper, I focus on identification and estimation of (reduced-form) response functions under random assignment of the treatment. By considering the "many small groups" case with an exogenous treatment, I can achieve point identification in a setting that has wide empirical applicability. On the other hand, randomization allows me to bypass the endogeneity issues that plague observational studies and focus on the structure of the response function, with emphasis on the restrictions that justify the different models that the literature has used in practice and their causal interpretation. Specifically, after defining a potential outcome $Y_{ig}(d, \mathbf{d}_g)$, I show how a general, nonparametric potential-outcome model can become a (reduced-form) LIM model under three conditions, namely (i) exchangeability, (ii) equal spillover effects under treatment and control status, and (iii) linear spillover effects. I also analyze the parameters that can be recovered by a misspecified LIM model.

## A1.2 Analysis of experiments with interference

By "analysis of experiments with interference" I refer to a body of research, developed primarily in statistics and epidemiology, that studies causal inference in experiments when the potential outcome of a unit can depend on the treatment assignments of other units (i.e., a failure of the SUTVA). Rubin (1990) and later Halloran and Struchiner (1991, 1995) extended the potential-outcomes causal framework by letting each unit's potential outcome to depend on the vector of treatment assignments in a sample. In this setting, the literature has mostly focused on four estimands. Given a sample with units $i = 1, \ldots, M$, the *direct effect* is the difference in potential outcomes for unit $i$ under treatment and control, given a vector of assignments for the remaining $M - 1$ units. The *indirect effect* is the difference in the outcomes of unit $i$, given own treatment assignment, for two possible assignments of the remaining $M - 1$ units. The *total effect* is the sum of the direct and indirect effects. Finally, the *overall effect* is the difference between the potential outcomes of unit $i$ under two alternative vector of treatment assignments. The corresponding average effects are defined

by simply averaging these estimands over the whole sample, as described below. As it is common in analysis of experiments (see e.g. Imbens and Rubin, 2015), potential outcomes are seen as fixed, and all the randomness comes through the treatment assignment mechanism.

The main complication that arises in a setting with interference is that the number of potential outcomes for each unit can become very large, taking up to $2^M$ values. Sobel (2006) introduced the assumption of *partial interference*, under which units in a sample are partitioned into groups, and interference can only occur between units in the same group. This assumption greatly simplifies the problem and seems to have been adopted by a vast majority of studies in this literature. Following the notation I defined in the paper, I will consider a sample of $g = 1, \ldots, G$ groups, each with $i = 1 \ldots, n_g + 1$ units.

Given the focus on finite populations with non-random outcomes, identification issues are largely absent from this literature, and interest is placed instead on finding unbiased estimators for the estimands of interest, estimating their variance and performing inference. Hudgens and Halloran (2008) discuss unbiased estimation and variance calculations under partial interference under two-stage randomization designs. They focus on finite-population versions of the estimands described above, in which individual potential outcomes are averaged over the distribution of the vector of neighbors' assignments. More precisely, given a probability distribution of treatment assignment parameterized by $\psi$, the individual average potential outcome under assignment $d$ is defined as:

$$\bar{Y}_{ig}(d, \psi) = \sum_{\mathbf{d}_g \in \mathcal{D}_g} Y_{ig}(d, \mathbf{d}_g) \mathbb{P}_\psi[\mathbf{D}_{(i)g} = \mathbf{d}_g | D_{ig} = d]$$

Based on this magnitude, the group average potential outcome and the (finite) population average potential outcome are given by:

$$\bar{Y}_g(d, \psi) = \frac{1}{n_g + 1} \sum_{i=1}^{n_g+1} \bar{Y}_{ig}(d, \psi), \qquad \bar{Y}(d, \psi) = \frac{1}{G} \sum_{g=1}^{G} \left( \frac{1}{n_g + 1} \sum_{i=1}^{n_g+1} \bar{Y}_{ig}(d, \psi) \right)$$

Then, the population average direct effect is given by $\bar{Y}(1, \psi) - \bar{Y}(0, \psi)$; given two parameterizations of the treatment assignment distribution, $\psi$ and $\phi$, the population average indirect effect is $\bar{Y}(0, \psi) - \bar{Y}(0, \phi)$; the population average total effect is $\bar{Y}(1, \psi) - \bar{Y}(0, \phi)$ (which is the sum of the direct and indirect effects). Hudgens and Halloran (2008) propose unbiased estimators for the above estimands, and provide variance estimators under exchangeability. Tchetgen Tchetgen and VanderWeele (2012) and Rigdon and Hudgens (2015) propose finite sample confidence intervals, while Liu and Hudgens (2014) study confidence intervals in a large-sample randomization inference context. Basse and Feller (2018) adapt the variance estimators to the case of varying group sizes and link the randomization inference framework with the regression framework.

The statistics literature focuses almost exclusively on inference for finite populations with fixed potential outcomes, in which all the randomness comes through the assignment

mechanism. The super-population approach, under which potential outcomes are drawn from a certain (infinite) population distribution, has two advantages over the finite population one. First, the parameters are defined with respect to a population of interest instead of a particular realization of a sample. In this sense, a super-population parameter has more external validity and more policy relevance than a magnitude which is only defined for a particular sample. Incidentally, some common population estimands of interest are not well defined when potential outcomes are fixed; for instance, the average treatment effect on the treated (ATT) is a random variable in a finite sample. Second, the population approach allows me to clearly distinguish the assumptions needed for identification from the ones needed for estimation and inference. This distinction gives a clearer conceptual picture of what conditions are required to identify the parameters of interest and what conditions are simplifications that permit estimating them and conducting hypothesis testing.

Among the few studies that analyze interference under a super-population approach, Philipson (2000) suggests conducting two-stage randomization experiments to analyze how average outcomes change in response to different shares of treated units in the population ("external effects"), while Baird, Bohren, McIntosh, and Özler (2018) perform power calculations for the above estimands (in its super-population version) under what they call saturation designs. Both studies consider a setting in which units are partitioned intro groups and potential outcomes are exchangeable.

Under both the finite-population and the super-population approaches, two-stage randomization has played a crucial role in the interference literature. In fact, some estimands of interest like the average indirect effect $\bar{Y}(0,\psi) - \bar{Y}(0,\phi)$ in Hudgens and Halloran (2008) are generally undefined under other designs like, for instance, simple randomization. Furthermore, the few studies discussing identification of population parameters tend to attribute identifying power to the 2SR design. For example, the abstract from Philipson (2000) states that "two-stage randomization schemes, which randomize allocation of treatments across communities and randomizes the treatments themselves within communities, are useful for identifying private and external treatment effects", while Baird, Bohren, McIntosh, and Özler (2018) claim: "[w]e show that [randomized saturation] designs identify a set of novel estimands" (page 2). However, I show in my paper that (i) spillover effects can be defined without reference to a specific assignment mechanism or experimental design, and (ii) these spillover effects can be identified as long as the assignment mechanism puts non-zero probability on each possible assignment. Specifically, I argue that a simple randomized (Bernoulli) experiment is enough to identify all the parameters of interest.

While most of this literature assumes partial interference, a recent body of research seeks to adapt the potential outcomes framework to more general structures of social interactions through arbitrary networks. When allowing for general interference, potential outcomes can depend on the treatment assignment of the whole population. In fact, the partial interference assumption can be seen precisely as a way to simplify this problem; in a networks setting, partial interference corresponds to the case with many independent networks (or alternatively,

a large network with a block diagonal adjacency matrix). Since estimation and inference can become infeasible when the structure of interactions is completely arbitrary, the main challenge faced by this literature is therefore to provide reasonable restrictions on the type of interference that can occur between units.

Some studies replace the partial interference by similar but more general restrictions on the spillovers structure. For instance, Leung (forthcoming) proposes restrictions on the structure of dependency graphs, which describe the correlation structure in a network, to perform asymptotic inference in a super-population framework. On the other hand, Eckles, Karrer, and Ugander (2017) study the bias of the global ATE estimator under different modeling assumptions and experimental designs. Choi (2017) considers identification under the assumption that the treatment effect is monotone.

# A2 Endogenous effects and structural models

As described in the literature review, a large literature in economics has studied identification in structural models in which the outcome of a unit depends on the outcomes of neighbors. While Manski (2013a), Manski (2013b) and Angrist (2014) have questioned the relevance of such models from a causal perspective, it is interesting to ask what type of structural model can justify the response functions that I study in this paper. To simplify the discussion, consider a setting with groups of size 2, that is, each unit has one neighbor. Suppose the potential outcomes $y_i$ are generated by the following system, where the arguments are suppressed to simplify notation:

$$y_1 = f(d_1, d_2, y_2, \varepsilon_1)$$
$$y_2 = f(d_2, d_1, y_1, \varepsilon_2)$$

This implies that

$$y_1 = f(d_1, d_2, f(d_2, d_1, y_1, \varepsilon_2), \varepsilon_1)$$
$$y_2 = f(d_2, d_1, f(d_1, d_2, y_2, \varepsilon_1), \varepsilon_2)$$

Depending on the form of the $f(\cdot, \cdot, \cdot)$, the above system may have one, zero or multiple equilibria. Suppose that $f(\cdot, \cdot, \cdot)$ is such that the system has a unique equilibrium. Then, the reduced form is given by:

$$y_i = \varphi(d_i, d_j, \varepsilon_i, \varepsilon_j)$$

Now, define $\phi(d_i, d_j, \varepsilon_i) = \mathbb{E}[\varphi(d_i, d_j, \varepsilon_i, \varepsilon_j)|\varepsilon_i]$, which integrates over $\varepsilon_j$.Then,

$$y_i = \varphi(d_i, d_j, \varepsilon_i) + u_{ij}$$

where $u_{ij} = \varphi(d_i, d_j, \varepsilon_i, \varepsilon_j) - \varphi(d_i, d_j, \varepsilon_i)$. Now, because $d_i$ and $d_j$ are binary, we have that:

$$\begin{aligned}\varphi(d_i, d_j, \varepsilon_i) &= \varphi_i^{00}(1-d_i)(1-d_j) + \varphi_i^{10} d_i(1-d_j) \\ &\quad + \varphi_i^{01}(1-d_i)d_j + \varphi_i^{11} d_i d_j \\ &:= \varphi_i^{00} + (\varphi_i^{10} - \varphi_i^{00})d_i + (\varphi_i^{01} - \varphi_i^{00})d_j \\ &\quad + (\varphi_i^{11} - \varphi_i^{01} - \varphi_i^{10} + \varphi_i^{00})d_i d_j\end{aligned}$$

and therefore

$$y_i = \varphi_i^{00} + (\varphi_i^{10} - \varphi_i^{00})d_i + (\varphi_i^{01} - \varphi_i^{00})d_j + (\varphi_i^{11} - \varphi_i^{01} - \varphi_i^{10} + \varphi_i^{00})d_i d_j + u_{ij}$$

The above expression can be relabeled to match the notation in the paper, ignoring the group subscript:

$$Y_i(d, d^1) = Y_{ij}(0) + \tau_i d + \gamma_i(0)d^1 + (\gamma_i(1) - \gamma_i(0))d \cdot d^1$$

where $Y_{ij,g}(0) = \phi_{ig}^{00} + u_{ij,g}$.

An important difference when allowing for endogenous effects is the presence of an additional term, $u_{ij}$, which depends on the heterogeneity of all units in the group. The presence of this term will generally introduce correlation between units in the same group. This feature does not affect identification, but has to be taken into account when performing inference.

Importantly,since the treatment indicators are binary, the reduced form can always be written in a fully saturated form, which does not require any assumptions on the structural equations, besides the restrictions that guarantee a unique equilibrium.

As an illustration, consider the following structural function:

$$y_i = f(d_i, y_j, \varepsilon_i) = \alpha_i + \beta_i d_i + \theta_i y_j + \delta_i d_i y_j$$

where $\alpha_i = \alpha(\varepsilon_i)$ and similarly for $\beta_i, \theta_i$ and $\delta_i$. Then,

$$\begin{cases} y_i = \alpha_i + \theta_i(\alpha_j + \theta_j y_i) & \text{if } d_i = 0,\ d_j = 0 \\ y_i = \alpha_i + \beta_i + (\theta_i + \delta_i)(\alpha_j + \theta_j y_i) & \text{if } d_i = 1,\ d_j = 0 \\ y_i = \alpha_i + \theta_i(\alpha_j + \beta_j + (\theta_j + \delta_j)y_i) & \text{if } d_i = 0,\ d_j = 1 \\ y_i = \alpha_i + \beta_i + (\theta_i + \delta_i)(\alpha_j + \beta_j + (\theta_j + \delta_j)y_i) & \text{if } d_i = 1,\ d_j = 1 \end{cases}$$

which implies the reduced form:

$$\begin{cases} y_i = \frac{\alpha_i+\theta_i\alpha_j}{1-\theta_i\theta_j} & \text{if } d_i=0,\, d_j=0 \\ y_i = \frac{\alpha_i+\theta_i\alpha_j}{1-(\theta_i+\delta_i)\theta_j} + \frac{\beta_i}{1-(\theta_i+\delta_i)\theta_j} & \text{if } d_i=1,\, d_j=0 \\ y_i = \frac{\alpha_i+\theta_i\alpha_j}{1-\theta_i(\theta_j+\delta_j)} + \frac{\beta_j\theta_i}{1-\theta_i(\theta_j+\delta_j)} & \text{if } d_i=0,\, d_j=1 \\ y_i = \frac{\alpha_i+\theta_i\alpha_j}{1-(\theta_i+\delta_i)(\theta_j+\delta_j)} + \frac{\beta_i}{1-(\theta_i+\delta_i)(\theta_j+\delta_j)} + \frac{\beta_j\theta_i}{1-(\theta_i+\delta_i(\theta_j+\delta_j)} & \text{if } d_i=1,\, d_j=1 \end{cases}$$

as long as $\theta_i\theta_j$, $(\theta_i+\delta_i)\theta_j$, $\theta_i(\theta_j+\delta_j)$ and $(\theta+\delta_i)(\theta_j+\delta_j)$ are different from 1 almost surely. This expression can be rewritten as before:

$$y_i = \varphi_i^{00} + (\varphi_i^{10}-\varphi_i^{00})d_i + (\varphi_i^{01}-\varphi_i^{00})d_j + (\varphi_i^{11}-\varphi_i^{01}-\varphi_i^{10}+\varphi_i^{00})d_id_j + u_{ij}$$

where now all the $\varphi_i$ terms are functions of the structural parameters. Notice that the neighbor's treatment assignment enters the reduced form in levels and with an interaction, even though it does not enter the structural equation. This is so because in the reduced form equation, $d_j$ captures the effect of $y_j$.

# A3 Unequally-sized groups

To explicitly account for different group sizes, let $n$ (the total number of peers in each group) take values in $\mathcal{N} = \{n_1, n_2, \ldots, n_K\}$ where $n_k \geq 1$ for all $k$ and $n_1 < n_2 < \ldots < n_K$. Let the potential outcome be $Y_{ig}(n, d, s(n))$ where $n \in \mathcal{N}$ and $s(n) \in \{0, 1, 2, \ldots, n\}$. Let $N_g$ be the observed value of $n_g$, $S_{ig}(n) = \sum_{j\neq i}^{n} D_{jg}$ and $S_{ig} = \sum_{k=1}^{K} S_{ig}(n_k)\mathbb{1}(N_g = n_k)$. The independence assumption can be modified to hold conditional on group size:

$$\{Y_{ig}(n, d, s(n)) : d = 0, 1, s(n) = 0, 1, \ldots, n\}_{i=1}^{n} \perp\!\!\!\perp \mathbf{D}_g(n))|N_g = n$$

where $\mathbf{D}_g(n)$ is the vector of all treatment assignments when the group size is $n+1$.

Under this assumption, we have that for $n \in \mathcal{N}$ and $s \leq n$,

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

The average observed outcome conditional on $N_g = n$ can be written as:

$$\begin{aligned} \mathbb{E}[Y_{ig}|D_{ig}, S_{ig}, N_g = n] = \mathbb{E}[Y_{ig}(n,0,0)] &+ \tau_0(n)D_{ig} \\ &+ \sum_{s=1}^{n} \theta_0(s,n)\mathbb{1}(S_{ig}=s)(1-D_{ig}) \\ &+ \sum_{s=1}^{n} \theta_1(s,n)\mathbb{1}(S_{ig}=s)D_{ig} \end{aligned}$$

The easiest approach is to simply run separate analyses for each group size and estimate

all the effects separately. In this case, it is possible to test whether spillover effects are different in groups with different sizes. The total number of parameters in this case is given by $\sum_{k=1}^{K}(n_k+1)$.

In practice, however, there may be cases in which group size has a rich support with only a few groups at each value $n_g$, so separate analyses may not be feasible. In such a setting, a possible solution is to impose an additivity assumption on group size. According to this assumption, the average direct and spillover effects do not change with group size. For example, the spillover effect of having one treated neighbor is the same in a group with two or three units. Under this assumption,

$$\begin{aligned}\mathbb{E}[Y_{ig}|D_{ig},S_{ig},N_g] = &\sum_{n_g\in\mathcal{N}_g}\alpha(n_g)\mathbb{1}(N_g=n_g)+\tau_0 D_{ig}\\ &+\sum_{s=1}^{N_g}\theta_0(s)\mathbb{1}(S_{ig}=s)(1-D_{ig})\\ &+\sum_{s=1}^{N_g}\theta_1(s)\mathbb{1}(S_{ig}=s)D_{ig}\end{aligned}$$

where the first sum can be seen in practice as adding group-size fixed effects. Then, the identification results and estimation strategies in the paper are valid after controlling for group-size fixed effects. Note that in this case the total number of parameters to estimate is $n_K+K-1$ where $n_K$ is the size of the largest group and $K$ is the total number of different group sizes.

Another possibility is to assume that for any constant $c\in\mathbb{N}$, $Y_{ig}(c\cdot n,d,c\cdot s)=Y_{ig}(n,d,s)$. This assumption allows us to rewrite the potential outcomes as a function of the ratio of treated peers, $Y_{ig}(d,s/n)$. Letting $P_{ig}=S_{ig}/N_g$, all the parameters can be estimated by running a regression including $D_{ig}$, $\mathbb{1}(P_{ig}=p)$ for all possible values of $p>0$ (excluding $p=0$ to avoid perfect collinearity) and interactions. In this case, the total number of parameters can be bounded by $n_1+\sum_{k=2}^{K}(n_k-1)$. Note that assuming that the potential outcomes depend only on the proportion of treated siblings does not justify in any way including the variable $P_{ig}$ linearly, as commonly done in linear-in-means models.

# A4 Including covariates

There are several reasons why one may want to include covariates when estimating direct and spillover effects. First, pre-treatment characteristics may help reduce the variability of the estimators and decrease small-sample bias, which is standard practice when analyzing randomly assigned programs. Covariates can also help get valid inference when the assignment mechanisms stratifies on baseline covariates. This can be done by simply augmenting Equation (1) with a vector of covariates $\boldsymbol{\gamma}'\mathbf{x}_{ig}$ which can vary at the unit or at the group level. The covariates can also be interacted with the treatment assignment indicators to explore

effect heterogeneity across observable characteristics (for example, by separately estimating effects for males and females.

Second, exogenous covariates can be used to relax the mean-independence assumption in observational studies. More precisely, if $\mathbf{X}_g$ is a matrix of covariates, a conditional mean-independence assumption would be

$$\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g, \mathbf{D}_g] = \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g]$$

which is a version of the standard unconfoundeness condition. The vector of covariates can include both individual-level and group-level characteristics.

Third, covariates can be included to make an exposure mapping more likely to be correctly specified. For instance, the exchangeability assumption can be relaxed by assuming it holds after conditioning on covariates, so that for any pair of treatment assignments $\mathbf{d}_g$ and $\tilde{\mathbf{d}}_g$ with the same number of ones,

$$\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g] = \mathbb{E}[Y_{ig}(d, \tilde{\mathbf{d}}_g)|\mathbf{X}_g]$$

As an example, exchangeability can be assumed to hold for all siblings with the same age, gender or going to the same school.

All the identification results in the paper can be adapted to hold after conditioning on covariates. In terms of implementation, when the covariates are discrete the parameters of interest can be estimated at each possible value of the matrix $\mathbf{X}_g$, although this strategy can worsen the dimensionality problem. Alternatively, covariates can be included in a regression framework after imposing parametric assumptions, for example, assuming the covariates enter linearly.

# A5 Additional empirical results

Table A1 shows the estimates for direct and spillover effects with the data used in the paper, but without imposing exchangeability. I define the order 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. Using these estimates, exchangeability is easily 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 results are given in the last two rows of the table, which is clearly not able to reject exchangeability in this case.

Table A1: Estimation results

| | coef | s.e. |
|---|---|---|
| $D_{ig}$ | 0.165** | 0.066 |
| $(1-D_{ig})D_{i1g}(1-D_{i2g})$ | 0.133** | 0.067 |
| $(1-D_{ig})(1-D_{i1g})D_{i2g})$ | 0.163** | 0.07 |
| $(1-D_{ig})D_{i1g}D_{i2g}$ | 0.14** | 0.056 |
| $D_{ig}D_{i1g}(1-D_{i2g})$ | -0.039 | 0.027 |
| $D_{ig}(1-D_{i1g})D_{i2g}$ | -0.044* | 0.026 |
| $D_{ig}D_{i1g}D_{i2g}$ | -0.051** | 0.025 |
| Constant | 0.706*** | 0.057 |
| Observations | | 504 |
| Chi-squared test | 0.444 | |
| p-value | 0.642 | |

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

# A6 Additional theoretical results

In what follows, let $\mathbf{0}_g$ and $\mathbf{1}_g$ be $n_g$-dimensional vectors of zeros and ones, respectively, let the true exposure mapping be $h_0(\cdot)$ and let $\mathbf{0} = h_0(\mathbf{0}_g)$.

**Lemma A1 (Difference in means)** *Under Assumption 1, the coefficient $\beta_{\mathsf{D}}$ from Equation (2) can be written as:*

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

**Lemma A2 (LIM regression)** *Under Assumption 1 and simple random assignment, the coefficient $\gamma_\ell$ from Equation (3) can be written as:*

$$\gamma_\ell = n_g \sum_{s=1}^{n_g} m(s) \left(\frac{s-\mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]}\right) \mathbb{P}[S_{ig}=s]$$

*where*

$$m(s) = \sum_{\mathbf{h}\in\mathcal{H}_0} \left(p\mathbb{E}[Y_{ig}(1,\mathbf{h}) - Y_{ig}(1,\mathbf{0})] + (1-p)\mathbb{E}[Y_{ig}(0,\mathbf{h}) - Y_{ig}(0,\mathbf{0})]\right) \mathbb{P}[\mathbf{H}_{ig}^0 = \mathbf{h}|S_{ig}=s].$$

**Lemma A3 (Interacted LIM regression)** *Under Assumption 1 and simple random assignment, the coefficient $\gamma_\ell^d$ from Equation (4) can be written as:*

$$\gamma_\ell^d = n_g \sum_{s=1}^{n_g} \tilde{m}(d,s) \left(\frac{s-\mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]}\right) \mathbb{P}[S_{ig}=s]$$

*where*

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

**Lemma A4 (Assignment probabilities)** *Let* $\hat{\pi}(\mathbf{a}) := \sum_g \sum_i \mathbb{1}_{ig}(\mathbf{a})/G(n+1)$. *Then under the assumptions of Lemma 5,*

$$\max_{\mathbf{a}\in\mathcal{A}_n} \left| \frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1 \right| \to_{\mathbb{P}} 0.$$

# A7 Proofs of additional results

**Proof of Lemma A1** By Lemma 1,

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

and thus

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

after adding and subtracting $\mathbb{E}[Y_{ig}(1,\mathbf{0})]$ and $\mathbb{E}[Y_{ig}(0,\mathbf{0})]$. □

**Proof of Lemma A2** By independence between $D_{ig}$ and $S_{ig}$ under simple random assignment,

$$\gamma_\ell = \frac{\mathbb{C}\mathrm{ov}(Y_{ig}, \bar{D}_g^{(i)})}{\mathbb{V}[\bar{D}_g^{(i)}]} = n_g \frac{\mathbb{C}\mathrm{ov}(Y_{ig}, S_{ig})}{\mathbb{V}[S_{ig}]} = n_g \frac{\mathbb{C}\mathrm{ov}(\mathbb{E}[Y_{ig}|S_{ig}], S_{ig})}{\mathbb{V}[S_{ig}]}$$

but $\mathbb{E}[Y_{ig}|S_{ig}] = p\mathbb{E}[Y_{ig}|D_{ig}=1, S_{ig}] + (1-p)\mathbb{E}[Y_{ig}|D_{ig}=0, S_{ig}]$, and

$$\mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}] = \sum_{s=0}^{n_g} \mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s]\mathbb{1}(S_{ig}=s).$$

and thus

$$\gamma_\ell = n_g \sum_{s=0}^{n_g} \mathbb{E}[Y_{ig}|D_{ig}=d, S_{ig}=s] \frac{\mathbb{C}\mathrm{ov}(\mathbb{1}(S_{ig}=s), S_{ig})}{\mathbb{V}[S_{ig}]}.$$

Next,

$$\begin{aligned}\mathbb{C}\text{ov}(\mathbb{1}(S_{ig}=s),S_{ig}) &= \mathbb{E}[\mathbb{1}(S_{ig}=s)S_{ig}]-\mathbb{E}[\mathbb{1}(S_{ig}=s)]\mathbb{E}[S_{ig}]\\ &= \mathbb{E}[S_{ig}|S_{ig}=s]\mathbb{P}[S_{ig}=s]-\mathbb{P}[S_{ig}=s]\mathbb{E}[S_{ig}]\\ &= (s-\mathbb{E}[S_{ig}])\mathbb{P}[S_{ig}]\end{aligned}$$

and thus

$$\gamma_\ell = n_g\sum_{s=0}^{n_g}(p\mu(1,s)+(1-p)\mu(0,s))\left(\frac{s-\mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]}\right)\mathbb{P}[S_{ig}].$$

where $\mu(d,s)=\mathbb{E}[Y_{ig}|D_{ig}=d,S_{ig}=s]$. Finally, from Lemma 1,

$$\begin{aligned}\mu(d,s) &= \sum_{\mathbf{h}\in\mathcal{H}_0}\mathbb{E}[Y_{ig}(d,\mathbf{h})]\mathbb{P}[H_{ig}^0=\mathbf{h}|D_{ig}=d,S_{ig}=s]\\ &= \sum_{\mathbf{h}\in\mathcal{H}_0}\mathbb{E}[Y_{ig}(d,\mathbf{h})]\mathbb{P}[H_{ig}^0=\mathbf{h}|S_{ig}=s]\\ &= \mathbb{E}[Y_{ig}(d,\mathbf{0})]+\sum_{\mathbf{h}\in\mathcal{H}_0}\left(\mathbb{E}[Y_{ig}(d,\mathbf{h})-Y_{ig}(d,\mathbf{0})]\right)\mathbb{P}[H_{ig}^0=\mathbf{h}|S_{ig}=s]\end{aligned}$$

where the second equality uses the fact that $D_{ig}$ and $\mathbf{D}_{(i)g}$ are independent under simple random assignment, and the result follows from the fact that

$$\sum_{s=0}^{n_g}\left(\frac{s-\mathbb{E}[S_{ig}]}{\mathbb{V}[S_{ig}]}\right)\mathbb{P}[S_{ig}]=0$$

and that $\mathbb{P}[H_{ig}^0=\mathbf{h}|S_{ig}=0]=\mathbb{1}(\mathbf{h}=\mathbf{0})$. □

**Proof of Lemma A3** This result follows from the same argument as the previous lemma but conditioning on $D_{ig}=d$. □

**Proof of Lemma A4** Take $\varepsilon>0$, then

$$\begin{aligned}\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\varepsilon\right] &\le \sum_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\varepsilon\right] \le |\mathcal{A}_n|\max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[|\hat{\pi}(\mathbf{a})-\pi(\mathbf{a})|>\varepsilon\pi(\mathbf{a})\right]\\ &= |\mathcal{A}_n|\max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[|N(\mathbf{a})-\mathbb{E}[N(\mathbf{a})]|>\varepsilon\mathbb{E}[N(\mathbf{a})]\right]\end{aligned}$$

Now,

$$N(\mathbf{a})-\mathbb{E}[N(\mathbf{a})]=\sum_g\sum_i\mathbb{1}_{ig}(\mathbf{a})-G(n+1)\pi(\mathbf{a})=\sum_g W_g$$

where $W_g = \sum_i \mathbb{1}_{ig}(\mathbf{a}) - (n+1)\pi(\mathbf{a}) = N_g(\mathbf{a}) - \mathbb{E}[N_g(\mathbf{a})]$. Note that the $W_g$ are independent and:

$$\begin{aligned}
\mathbb{E}[W_g] &= 0 \\
|W_g| &\leq (n+1)\max\{\pi(\mathbf{a}), 1-\pi(\mathbf{a})\} \\
\mathbb{V}[W_g] &= \mathbb{V}\left[\sum_i \mathbb{1}_{ig}(\mathbf{a})\right] = \sum_i \mathbb{V}[\mathbb{1}_{ig}(\mathbf{a})] + 2\sum_i \sum_{j>i} \mathbb{C}\mathrm{ov}(\mathbb{1}_{ig}(\mathbf{a}), \mathbb{1}_{jg}(\mathbf{a})) \\
&= (n+1)\pi(\mathbf{a})(1-\pi(\mathbf{a})) + (n+1)(n+2)\{\mathbb{E}[\mathbb{1}_{ig}(\mathbf{a})\mathbb{1}_{jg(\mathbf{a})}] - \pi(\mathbf{a})^2\} \\
&\leq (n+1)\pi(\mathbf{a})(1-\pi(\mathbf{a})) + (n+1)(n+2)\pi(\mathbf{a})(1-\pi(\mathbf{a})) \\
&= (n+1)(n+3)\pi(\mathbf{a})(1-\pi(\mathbf{a}))
\end{aligned}$$

Then, by Bernstein's inequality,

$$\begin{aligned}
\mathbb{P}\left[|W_g| > \varepsilon\mathbb{E}[N(\mathbf{a})]\right] &\leq 2\exp\left\{-\frac{\mathbb{E}[N(\mathbf{a})]^2\varepsilon^2}{\sum_g \mathbb{V}[W_g] + \frac{1}{3}(n+1)\max\{\pi(\mathbf{a}), 1-\pi(\mathbf{a})\}\mathbb{E}[N\mathbf{a})]\varepsilon}\right\} \\
&= 2\exp\left\{-\frac{\frac{1}{2}G^2(n+1)^2\pi(\mathbf{a})^2\varepsilon^2}{G(n+1)(n+3)\pi(\mathbf{a})(1-\pi(\mathbf{a})) + \frac{1}{3}G(n+1)^2\pi(\mathbf{a})\max\{\pi(\mathbf{a}), 1-\pi(\mathbf{a})\}\varepsilon}\right\} \\
&= 2\exp\left\{-\frac{\frac{1}{2}G\pi(\mathbf{a})\varepsilon^2}{\frac{n+3}{n+1}(1-\pi(\mathbf{a})) + \frac{1}{3}\max\{\pi(\mathbf{a}), 1-\pi(\mathbf{a})\varepsilon\}}\right\} \\
&\leq 2\exp\left\{-\frac{\frac{1}{2}G\pi(\mathbf{a})\varepsilon^2}{\frac{n+3}{n+1} + \frac{\varepsilon}{3}}\right\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \varepsilon\right] &\leq |\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[|N(\mathbf{a}) - \mathbb{E}[N(\mathbf{a})]| > \varepsilon\mathbb{E}[N(\mathbf{a})]\right] \\
&\leq 2\exp\left\{-G\underline{\pi}_n\left(\frac{\frac{1}{2}\varepsilon^2}{\frac{n+3}{n+1} + \frac{\varepsilon}{3}} - \frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n}\right)\right\} \to 0.
\end{aligned}$$

as required. □

# A8 Proofs of main results

**Proof of Lemma 1** If $\mathbb{P}[D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h}]$,

$$\begin{aligned}\mathbb{E}[Y_{ig}|D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h}] &= \sum_{\mathbf{h}_0}\mathbb{E}[Y_{ig}|D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h},\mathbf{H}^0_{ig}=\mathbf{h}_0]\\ &\quad\times\mathbb{P}[\mathbf{H}^0_{ig}=\mathbf{h}_0|D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h}]\\ &= \sum_{\mathbf{h}_0}\mathbb{E}[Y_{ig}(d,\mathbf{h}_0)|D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h},\mathbf{H}^0_{ig}=\mathbf{h}_0]\\ &\quad\times\mathbb{P}[\mathbf{H}^0_{ig}=\mathbf{h}^0|D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h}]\\ &= \sum_{\mathbf{h}_0}\mathbb{E}[Y_{ig}(d,\mathbf{h}_0)]\mathbb{P}[\mathbf{H}^0_{ig}=\mathbf{h}_0|D_{ig}=d,\mathbf{H}_{ig}=\mathbf{h}]\end{aligned}$$

where the first equality follows from the law if iterated expectations, the second equality follows by definition of the observed outcomes and the third equality follows from random assignment of the treatment vector given that both $\mathbf{H}^0_{ig}$ and $\mathbf{H}_{ig}$ are deterministic functions of $\mathbf{D}_g$. Finally, if $h_0(\cdot)$ is coarser than $h(\cdot)$, then $\mathbf{H}_{ig}=\mathbf{h}$ uniquely determines the value of $\mathbf{H}^0_{ig}$ and the result follows. □

**Proof of Lemma 2** Follows from Lemma A1 setting $h_0(\mathbf{d}_g)=\mathbf{1}_g'\mathbf{d}_g$. □

**Proof of Lemma 3** Follows from Lemma A2 setting $h_0(\mathbf{d}_g)=\mathbf{1}_g'\mathbf{d}_g$. □

**Proof of Lemma 4** Follows from Lemma A3 setting $h_0(\mathbf{d}_g)=\mathbf{1}_g'\mathbf{d}_g$. □

**Proof of Lemma 5** Take a constant $c\in\mathbb{R}$. Then

$$\mathbb{P}\left[\min_{\mathbf{a}\in\mathcal{A}_n}N(\mathbf{a})\leq c\right]\leq|\mathcal{A}_n|\max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}[N(\mathbf{a})\leq c].$$

Now, for any $\delta>0$,

$$\begin{aligned}\mathbb{P}[N(\mathbf{a})\leq c] &= \mathbb{P}\left[N(\mathbf{a})\leq c,\left|\frac{\hat\pi(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\delta\right]+\mathbb{P}\left[N(\mathbf{a})\leq c,\left|\frac{\hat\pi(\mathbf{a})}{\pi(\mathbf{a})}-1\right|\leq\delta\right]\\ &\leq \mathbb{P}\left[\left|\frac{\hat\pi(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\delta\right]+\mathbb{P}[N(\mathbf{a})\leq c,G(n+1)\pi(\mathbf{a})(1-\delta)\leq N(\mathbf{a})\leq\pi(\mathbf{a})G(n+1)(1+\delta)]\\ &\leq \mathbb{P}\left[\left|\frac{\hat\pi(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\delta\right]+\mathbb{1}(G(n+1)\pi(\mathbf{a})\leq c/(1-\delta))\\ &\leq \mathbb{P}\left[\left|\frac{\hat\pi(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\delta\right]+\mathbb{1}(G(n+1)\underline{\pi}_n\leq c/(1-\delta))\end{aligned}$$

which implies

$$|\mathcal{A}_n|\max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}[N(\mathbf{a})\leq c]\leq|\mathcal{A}_n|\max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[\left|\frac{\hat\pi(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\delta\right]+|\mathcal{A}_n|\mathbb{1}(G(n+1)\underline{\pi}_n\leq c/(1-\delta))$$

which converges to zero under condition 5 and using Lemma A4. □

**Proof of Theorem 1** All the estimators below are only defined when $\mathbb{1}(N(\mathbf{a}) > 0)$. Because under the conditions for Lemma 5 this event occurs with probability approaching one, the indicator will be omitted to simplify the notation. Let $\varepsilon_{ig}(\mathbf{a}) = Y_{ig} - \mathbb{E}[Y_{ig}|\mathbf{A}_{ig} = \mathbf{a}]$. For the consistency part, we have that

$$\frac{\sum_g \sum_i \varepsilon_{ig}(\mathbf{a})\mathbb{1}_{ig(\mathbf{a})}}{N(\mathbf{a})} = \frac{\sum_g \sum_i (\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}| > \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}| > \xi_n)])\mathbb{1}_{ig(\mathbf{a})}}{N(\mathbf{a})} + \frac{\sum_g \sum_i (\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}| \leq \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}| \leq \xi_n)])\mathbb{1}_{ig(\mathbf{a})}}{N(\mathbf{a})}$$

for some increasing sequence of constants $\xi_n$ whose rate will be determined along the proof. Let

$$\underline{\varepsilon}_{ig}(\mathbf{a}) = \varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| \leq \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| \leq \xi_n)]$$

and

$$\bar{\varepsilon}_{ig}(\mathbf{a}) = \varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| > \xi_n) - \mathbb{E}[\varepsilon_{ig}(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| > \xi_n)]$$

For the first term,

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\right| > Mr_n \middle| \mathbf{A}\right] \leq |\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[\left|\frac{\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\right| > Mr_n \middle| \mathbf{A}\right]$$

For the right-hand side, by Bernstein's inequality

$$\begin{aligned}\mathbb{P}\left[\left|\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a})Mr_n \middle| \mathbf{A}\right] &\leq 2\exp\left\{-\frac{1}{2}\frac{M^2 r_n^2 N(\mathbf{a})^2}{\sigma^2(\mathbf{a})N(\mathbf{a}) + 2\xi_n M r_n N(\mathbf{a})/3}\right\} \\ &= 2\exp\left\{-\frac{1}{2}\frac{M^2 r_n^2 N(\mathbf{a})}{\sigma^2(\mathbf{a}) + 2M\xi_n r_n/3}\right\} \\ &\leq 2\exp\left\{-\frac{1}{2}\frac{M^2 r_n^2 \min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a})}{\bar{\sigma}^2 + 2M\xi_n r_n/3}\right\}\end{aligned}$$

Set

$$r_n = \sqrt{\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}}$$

Next, use the fact that

$$\frac{\min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a})}{G(n+1)\underline{\pi}_n} \to_{\mathbb{P}} 1$$

which follows from Lemma A4, since

$$\mathbb{P}[\pi(\mathbf{a})(1-\varepsilon) \leq \hat{\pi}(\mathbf{a}) \leq \pi(\mathbf{a})(1+\varepsilon),\ \forall\, \mathbf{a}] \to 1$$

for any $\varepsilon > 0$ and

$$\begin{aligned}\mathbb{P}[\pi(\mathbf{a})(1-\varepsilon) \leq \hat{\pi}(\mathbf{a}) \leq \pi(\mathbf{a})(1+\varepsilon),\, \forall\, \mathbf{a}] &\leq \mathbb{P}[\underline{\pi}_n(1-\varepsilon) \leq \min_{\mathbf{a}} \hat{\pi}(\mathbf{a}) \leq \underline{\pi}_n(1+\varepsilon)] \\ &= \mathbb{P}\left[\left|\frac{\min_{\mathbf{a}} \hat{\pi}(\mathbf{a})}{\underline{\pi}_n} - 1\right| \leq \varepsilon\right].\end{aligned}$$

Then,

$$\mathbb{P}\left[\left|\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a}) \mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a}) M r_n \middle| \mathbf{A}\right] \leq 2\exp\left\{-\frac{1}{2}\frac{M^2 \log|\mathcal{A}_n|(1+o_{\mathbb{P}}(1))}{\bar{\sigma}^2 + 2M\xi_n r_n/3}\right\}$$

and therefore

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\sum_g \sum_i \underline{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}\right| > M r_n \middle| \mathbf{A}\right] \leq 2\exp\left\{\log|\mathcal{A}_n|\left(1 - \frac{1}{2}\frac{M^2(1+o_{\mathbb{P}}(1))}{\bar{\sigma}^2 + 2M r_n \xi_n/3}\right)\right\}$$

which can be made arbitrarily small for sufficiently large $M$ as long as $r_n \xi_n = O(1)$.

For the second term, by Markov's inequality

$$\begin{aligned}\mathbb{P}\left[\left|\sum_g \sum_i \bar{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a}) M r_n \middle| \mathbf{A}\right] &\leq \frac{\mathbb{E}[\varepsilon_{ig}^2(\mathbf{a})\mathbb{1}(|\varepsilon_{ig}(\mathbf{a})| > \xi_n)] N(\mathbf{a})}{M^2 r_n^2 N(\mathbf{a})^2} \\ &\leq \frac{b^{2+\delta}}{M^2 \xi_n^{\delta}} \frac{1}{r_n^2 \min_{\mathbf{a}\in\mathcal{A}_n} N(\mathbf{a})} \\ &= \frac{b^{2+\delta}}{M^2} \frac{1}{\xi_n^{\delta} \log|\mathcal{A}_n|(1+o_{\mathbb{P}}(1))}\end{aligned}$$

so that

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\sum_g \sum_i \bar{\varepsilon}_{ig}(\mathbf{a})\mathbb{1}_{ig}(\mathbf{a})\right| > N(\mathbf{a}) M r_n \middle| \mathbf{A}\right] \leq \frac{b^{2+\delta}}{M^2} \frac{|\mathcal{A}_n|}{\xi_n^{\delta} \log|\mathcal{A}_n|(1+o_{\mathbb{P}}(1))}$$

Finally, set $\xi_n = r_n^{-1}$. Then, the above term can be made arbitrarily small for $M$ sufficiently large, as long as

$$\frac{|\mathcal{A}_n|}{\log|\mathcal{A}_n|}\left(\frac{\log|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n}\right)^{\delta/2} = O(1)$$

Setting $\delta = 2$, this condition reduces to:

$$\frac{|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n} = O(1)$$

Therefore,

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

The proof for the standard error estimator uses the same reasoning after replacing $\varepsilon_{ig}(\mathbf{a})$ by $\hat{\varepsilon}_{ig}^2(\mathbf{a})$ and using consistency of $\hat{\mu}(\mathbf{a})$.

For the second part, we want to bound

$$\Delta = \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[ \frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \right] - \Phi(x) \right|$$

$$\begin{aligned}
\Delta &= \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[ \frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \right] - \Phi(x) \right| \\
&= \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{E}\left\{ \mathbb{P}\left[ \frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right\} \right| \\
&\leq \mathbb{E}\left\{ \max_{\mathbf{a}\in\mathcal{A}_\mathbf{n}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[ \frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| \right\}
\end{aligned}$$

Then,

$$\left| \mathbb{P}\left[ \frac{\hat{\mu}(\mathbf{a}) - \mu(\mathbf{a})}{\sqrt{\mathbb{V}[\hat{\mu}(\mathbf{a})|\mathbf{A}]}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| = \left| \mathbb{P}\left[ \frac{\sum_g \sum_i \varepsilon_{ig} \mathbb{1}_{ig}(\mathbf{a})}{\sigma(\mathbf{a})\sqrt{N(\mathbf{a})}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right|$$

By the Berry-Esseen bound,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[ \frac{\sum_g \sum_i \varepsilon_{ig} \mathbb{1}_{ig}(\mathbf{a})}{\sigma(\mathbf{a})\sqrt{N(\mathbf{a})}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| \leq \frac{Cb^3}{\underline{\sigma}^3} \cdot \frac{1}{\sqrt{N(\mathbf{a})}}$$

But

$$\frac{1}{N(\mathbf{a})} = O_\mathbb{P}\left( \frac{1}{G(n+1)\pi(\mathbf{a})} \right)$$

Therefore,

$$\max_{\mathbf{a}\in\mathcal{A}_n} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[ \frac{\sum_g \sum_i \varepsilon_{ig} \mathbb{1}_{ig}(\mathbf{a})}{\sigma(\mathbf{a})\sqrt{N(\mathbf{a})}} \leq x \middle| \mathbf{A} \right] - \Phi(x) \right| \leq \frac{Cb^3}{\underline{\sigma}^3} \cdot O_\mathbb{P}\left( \frac{1}{\sqrt{G(n+1)\underline{\pi}_n}} \right)$$

and the result follows. □

**Proof of Theorem 2** We want to bound:

$$\Delta^*(\mathbf{a}) = \sup_x \left| \mathbb{P}^*\left[ \frac{\hat{\mu}^*(\mathbf{a}) - \hat{\mu}(\mathbf{a})}{\sqrt{\mathbb{V}^*[\hat{\mu}(\mathbf{a})]}} \leq x \right] - \Phi(x) \right|$$

uniformly over $\mathbf{a}$, where

$$\hat{\mu}^*(\mathbf{a}) = \frac{\sum_g \sum_i Y_{ig}^* \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})}$$

if the denominator is non-zero, and zero otherwise, and where

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

Then, if $N(\mathbf{a}) > 0$,

$$\begin{aligned}
\mathbb{E}^*[\hat{\mu}^*(\mathbf{a})] &= \hat{\mu}(\mathbf{a}) \\
\mathbb{V}^*[\hat{\mu}^*(\mathbf{a})] &= \frac{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})^2}
\end{aligned}$$

The centered and scaled statistic is given by:

$$\frac{\sum_g \sum_i \hat{\varepsilon}_{ig} \mathbb{1}_{ig}(\mathbf{a}) w_{ig}}{\sqrt{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}}$$

By Berry-Esseen,

$$\sup_x \left| \mathbb{P}^* \left[ \frac{\sum_g \sum_i \hat{\varepsilon}_{ig} \mathbb{1}_{ig}(\mathbf{a}) w_{ig}}{\sqrt{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}} \leq x \right] - \Phi(x) \right| \leq C \frac{\sum_g \sum_i |\hat{\varepsilon}_{ig}|^3 \mathbb{1}_{ig}(\mathbf{a})/N(\mathbf{a})}{\left( \sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})/N(\mathbf{a}) \right)^{3/2}} \cdot \frac{1}{\sqrt{N(\mathbf{a})}}$$

We also have that

$$\begin{aligned}
\frac{\sum_g \sum_i |\hat{\varepsilon}_{ig}|^3 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} &\leq \frac{\sum_g \sum_i |Y_{ig} - \mu(\mathbf{a})|^3 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} + |\bar{Y}(\mathbf{a}) - \mu(\mathbf{a})|^3 + O_{\mathbb{P}}(N(\mathbf{a})^{-2}) \\
&= \mathbb{E}[|Y_{ig} - \mu(\mathbf{a})|^3] + O_{\mathbb{P}}(N(\mathbf{a})^{-1})
\end{aligned}$$

and

$$\begin{aligned}
\frac{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} &= \frac{\sum_g \sum_i (Y_{ig} - \mu(\mathbf{a}))^2 \mathbb{1}_{ig}(\mathbf{a})}{N(\mathbf{a})} + (\bar{Y}(\mathbf{a}) - \mu(\mathbf{a}))^2 \\
&= \sigma^2(\mathbf{a}) + O_{\mathbb{P}}(N(\mathbf{a})^{-1})
\end{aligned}$$

Then,

$$\begin{aligned}
\Delta^*(\mathbf{a}) &\leq \sup_x \left| \mathbb{P}^* \left[ \frac{\sum_g \sum_i \hat{\varepsilon}_{ig} \mathbb{1}_{ig}(\mathbf{a}) w_{ig}}{\sqrt{\sum_g \sum_i \hat{\varepsilon}_{ig}^2 \mathbb{1}_{ig}(\mathbf{a})}} \leq x \right] - \Phi(x) \right| \mathbb{1}(N(\mathbf{a}) > 0) + 2\mathbb{1}(N(\mathbf{a}) = 0) \\
&= C \frac{\mathbb{E}[|Y_{ig} - \mu(\mathbf{a})|^3] + O_{\mathbb{P}}(N(\mathbf{a})^{-1})}{[\sigma^2(\mathbf{a}) + O_{\mathbb{P}}(N(\mathbf{a})^{-1}))]^{3/2}} \cdot \frac{\mathbb{1}(N(\mathbf{a}) > 0)}{\sqrt{N(\mathbf{a})}} + 2\mathbb{1}(N(\mathbf{a}) = 0)
\end{aligned}$$

and the result follows from the facts that $\mathbb{P}[\min_{\mathbf{a}} N(\mathbf{a}) = 0] \to 0$ and by Lemma A4. □

**Proof of Corollary 1** Under exchangeability $\pi(\mathbf{a}) = \pi(d, s) = p^d(1-p)^{1-d}\binom{n}{s}p^s(1-p)^{n-s} = \binom{n}{s}p^{s+d}(1-p)^{n+1-s-d}$. This function is minimized at $\underline{\pi}_n = \underline{p}^{n+1} \propto \underline{p}^n$ where

$\underline{p} = \min\{p, 1-p\}$. Thus,

$$\frac{\log|\mathcal{A}_n|}{G\underline{p}^n} = \exp\left\{-\log G\left(1 - \frac{n+1}{\log G}2\log\underline{p} - \frac{\log\log|\mathcal{A}_n|}{\log G}\right)\right\}$$

and since $|\mathcal{A}_n| = 2(n+1)$, this term converges to zero when $(n+1)/\log G \to 0$. □

**Proof of Corollary 2** Under exchangeability, $\pi(\mathbf{a}) = \pi(d,s) = q_{d+s}\left(\frac{s+1}{n+1}\right)^d \times \left(1 - \frac{s}{n+1}\right)^{1-d}$ which in this case equals $\frac{1}{n+1}\left(\frac{s+1}{n+1}\right)^d\left(1 - \frac{s}{n+1}\right)^{1-d}$. This function has two minima, one at $(d,s) = (0,n)$ and one at $(d,s) = (1,0)$, giving the same minimized value of $\underline{\pi}_n = (n+1)^{-2}$. Then,

$$\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n} = \exp\left\{-\log G\left(1 - \frac{\log(n+1)}{\log G}4 - \frac{\log\log 2(n+1)}{\log G} + o(1)\right)\right\} \to 0$$

if $\log(n+1)/\log G \to 0$. □

# References


Angrist, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108.

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.

Basse, G., and A. Feller (2018): "Analyzing two-stage experiments in the presence of interference," *Journal of the American Statistical Association*, 113(521), 41–55.

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.

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

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150(1), 41 – 55.

Brock, W. A., and S. N. Durlauf (2001): "Discrete Choice with Social Interactions," *The Review of Economic Studies*, 68(2), 235–260.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009): "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27(3), 439–464.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81(3), 855–882.

Choi, D. (2017): "Estimation of Monotone Treatment Effects in Network Experiments," *Journal of the American Statistical Association*, 112(519), 1147–1155.

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.

Dieye, R., H. Djebbari, and F. Barrera-Osorio (2014): "Accounting for Peer Effects in Treatment Response," *IZA Discussion Paper*, (8340).

Eckles, D., B. Karrer, and J. Ugander (2017): "Design and Analysis of Experiments in Networks: Reducing Bias from Interference," *Journal of Causal Inference*, 5(1).

Graham, B. S. (2008): "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, 76(3), 643–660.

Graham, B. S., and J. Hahn (2005): "Identification and estimation of the linear-in-means model of social interactions," *Economics Letters*, 88(1), 1–6.

Halloran, M. E., and C. J. Struchiner (1991): "Study Designs for Dependent Happenings," *Epidemiology*, 2(5), 331–338.

——— (1995): "Causal Inference in Infectious Diseases," *Epidemiology*, 6(2), 142–151.

Hoxby, C. M., and G. Weingarth (2005): "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," *Working paper*.

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. 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.

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

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.

Leung, M. (forthcoming): "Treatment and Spillover Effects under Network Interference," *The Review of Economics and Statistics*.

Liu, L., and M. G. Hudgens (2014): "Large Sample Randomization Inference of Causal Effects in the Presence of Interference," *Journal of the American Statistical Association*, 109(505), 288–301.

Lyle, D. S. (2007): "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *The Review of Economics and Statistics*, 89(2), 289–299.

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.

Philipson, T. J. (2000): "External Treatment Effects and Program Implementation Bias," *NBER Technical Working Papers*, (250).

Rigdon, J., and M. G. Hudgens (2015): "Exact confidence intervals in the presence of interference," *Statistics & Probability Letters*, 105(6), 130–135.

Rubin, D. B. (1990): "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science*, 5(4), 472–480.

Sacerdote, B. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 116(2), 681–704.

——— (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," in *Handbook of the Economics of Education*, ed. by S. M. Eric A. Hanushek, and L. Woessmann, vol. 3, pp. 249 – 277.

——— (2014): "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?," *Annual Review of Economics*, 6(1), 253–272.

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.

Stinebrickner, R., and T. R. Stinebrickner (2006): "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds," *Journal of Public Economics*, 90(8), 1435–1454.

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.

Zimmerman, D. J. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, 85(1), 9–23.