Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Randomization Inference for Peer Effects

Xinran Li, Peng Ding, Qian Lin, Dawei Yang & Jun S. Liu

To cite this article: Xinran Li, Peng Ding, Qian Lin, Dawei Yang & Jun S. Liu (2019) Randomization Inference for Peer Effects, Journal of the American Statistical Association, 114:528, 1651-1664, DOI: 10.1080/01621459.2018.1512863

To link to this article: https://doi.org/10.1080/01621459.2018.1512863

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 114, NO. 528, 1651–1664: Theory and Methods
https://doi.org/10.1080/01621459.2018.1512863

Randomization Inference for Peer Effects

Xinran Li[a], Peng Ding[b], Qian Lin[c], Dawei Yang[d], and Jun S. Liu[a]

[a]Department of Statistics, Harvard University, Cambridge, MA; [b]Department of Statistics, University of California, Berkeley, CA; [c]Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing, P. R. China; [d]Bureau of Personnel of Chinese Academy of Sciences & School of Education of Peking University, Beijing, P. R. China

ABSTRACT

Many previous causal inference studies require no interference, that is, the potential outcomes of a unit do not depend on the treatments of other units. However, this no-interference assumption becomes unreasonable when a unit interacts with other units in the same group or cluster. In a motivating application, a top Chinese university admits students through two channels: the college entrance exam (also known as Gaokao) and recommendation (often based on Olympiads in various subjects). The university randomly assigns students to dorms, each of which hosts four students. Students within the same dorm live together and have extensive interactions. Therefore, it is likely that peer effects exist and the no-interference assumption does not hold. It is important to understand peer effects, because they give useful guidance for future roommate assignment to improve the performance of students. We define peer effects using potential outcomes. We then propose a randomization-based inference framework to study peer effects with arbitrary numbers of peers and peer types. Our inferential procedure does not assume any parametric model on the outcome distribution. Our analysis gives useful practical guidance for policy makers of the university. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received January 2017
Revised April 2018

KEYWORDS
Causal inference; Design-based inference; Grade point average (GPA); Interference; Optimal treatment assignment; Spillover effect.

1. Introduction

1.1. Causal Inference, Interference, and Peer Effects

The classical potential outcomes framework (Neyman 1923) assumes no interference among experimental units (Cox 1958), that is, the potential outcomes of a unit are functions of its own treatment but not others' treatments. This constitutes an important part of Rubin (1980)'s Stable Unit Treatment Value Assumption (SUTVA). In some experiments, interference is a nuisance, and researchers try to avoid it by isolating units. Interference, however, is unavoidable in many studies when units have interactions with each other. Examples include vaccine trials for infectious diseases in epidemiology (Halloran and Struchiner 1991, 1995; Perez-Heydrich et al. 2014), group-randomized trials in education (Hong and Raudenbush 2006; Vanderweele et al. 2013), and interventions on networks in sociology (An 2011; VanderWeele and An 2013), political science (Nickerson 2008; Ichino and Schündeln 2012; Bowers, Fredrickson, and Panagopoulos 2013), and economics (Manski 1993; Sacerdote 2001; Miguel and Kremer 2004; Graham, Imbens, and Ridder 2010; Goldsmith-Pinkham and Imbens 2013; Arpino and Mattei 2016). Ogburn and VanderWeele (2014) discussed different types of interference. Forastiere, Airoldi, and Mealli (2016) showed that ignoring interference can lead to biased inferences. It is important to study the pattern of interference in some applications, because it is of scientific interest and useful for decision-making. For example, Sacerdote (2001) found significant peer effects in student outcomes (e.g., GPA and fraternity membership) among students living in the same dorm of Dartmouth College. Based on this, Bhattacharya (2009) discussed the optimal peer assignment.

1.2. Motivating Application in Education

Our motivation comes from a dataset of a top Chinese university. It contains a rich set of variables of the students: family background, the channels through which they were admitted, roommates' information, GPAs, etc.

The university admits students through two primary channels: the college entrance exam (also known as Gaokao) and recommendation. Gaokao is an annual test in China to assess students' knowledge in various subjects. Every university has its own minimal test score threshold to admit students. Students from Gaokao study all subjects and often have broader knowledge. Students from recommendation do not need to take Gaokao. They win awards in national or international Olympiads in mathematics, physics, chemistry, biology, or informatics. They concentrate on a certain subject for the corresponding Olympiad. They may even take some college courses on that subject during their high school years. Most of them choose majors related to the subject they focused on in high schools. Overall, students admitted through these two channels have different training and thus different attributes. Students from recommendation generally perform better in GPAs than students from Gaokao.

After entering the university, students usually live in four-person rooms for four years. They often study together and interact with each other. We know that two types of students,

CONTACT Peng Ding pengdingpku@berkeley.edu Department of Statistics, University of California, Berkeley, CA 94720.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

from Gaokao and recommendation, have different training in high schools. It is then natural to ask the following questions. Is it beneficial for students from Gaokao to live with students from recommendation, or vice versa? Is there an optimal combination of roommate types for the performance of a certain student? Is there an optimal roommate assignment to maximize the performance of all students? These questions are all about peer effects among students.

1.3. Literature Review and Our Contribution

With interference, the potential outcomes of a unit can depend on its own treatment and others' treatments in various ways. Therefore, causal inference with interference has different mathematical forms. Like many other causal inference problems, there are at least two inferential frameworks for causal inference with interference: the Fisherian and Neymanian perspectives. Under the Fisherian view, Rosenbaum (2007), Luo et al. (2012), Aronow (2012), Bowers, Fredrickson, and Panagopoulos (2013), Rigdon and Hudgens (2015), Athey, Eckles, and Imbens (2018), and Basse, Feller, and Toulis (2019) proposed exact randomization tests for detecting causal effects with interference, and constructed confidence intervals for certain causal parameters by inverting tests. Choi (2017) discussed a related approach under the monotone treatment effect assumption. Under the Neymanian view, Hudgens and Halloran (2008) discussed point and interval estimation for several causal estimands with interference under two-stage randomized experiments on both the group and individual levels. Liu and Hudgens (2014) then established the large sample theory for these estimators. Aronow and Samii (2017), Basse and Feller (2018), and Sävje, Aronow, and Hudgens (2017) extended the discussion to other general contexts. The Fisherian and Neymanian views are both randomization based in the sense that the uncertainty in testing or estimation comes solely from the treatment assignment mechanism, and all the potential outcomes are fixed constants. When two-stage randomization is infeasible, we need certain unconfoundedness assumptions to ensure identifiability of causal effects. Tchetgen and VanderWeele (2012) proposed an inverse probability weighting estimator. Perez-Heydrich et al. (2014) applied this methodology to assess effects of cholera vaccination. Liu, Hudgens, and Becker-Dreps (2016) studied the theoretical properties. Other studies (Sacerdote 2001; Toulis and Kao 2013) relied on parametric modeling assumptions.

Our framework for peer effects furthers the literature in several ways. First, we define peer effects using potential outcomes. Unlike in previous work (e.g., Sobel 2006; Hudgens and Halloran 2008), our estimands do not involve averages over the treatment assignment. We separate the causal estimands from the treatment assignment. As Rubin (2005) argued, the former are functions of the potential outcomes, and the latter induces randomness and governs the statistical inference. Second, previous works discussed external interventions with known networks, clusters, or groups. Our hypothetical intervention is the roommate assignment in the motivating application. It forms a "network" among units, which further causes interference and peer effects. We explain the distinction between the two types of interference in detail in Section 2.5. Our setting is similar to Sacerdote (2001)'s. However, we formalize the problem using potential outcomes instead of linear models and allow for causal interpretations without imposing model assumptions. Third, we propose randomization-based point estimators, prove their asymptotic Normalities, and construct confidence intervals. We further derive the optimal roommate assignment to maximize the performance of students. The inferential framework is Neymanian, similar to those of Hudgens and Halloran (2008) and Aronow and Samii (2017). Fourth, we apply the new method to the dataset from a top Chinese university and find important policy implications. We relegate all the technical details to the supplementary material.

2. Notation and Framework for Peer Effects

2.1. Potential Outcomes with Peers

We consider an experiment with $n = m(K+1)$ units, where m is the number of groups and $K+1$ is the size of each group. Each unit has K peers in the same group. The group and peers correspond to room and roommates in our motivating application, where $K = 3$ is the number of roommates for each student. Let Z_i be the treatment assignment for unit i, which is a set consisting of the identity numbers of his/her K peers, that is, $Z_i = \{j : \text{units } j \text{ and } i \text{ are in the same group}\}$. In the motivating application, Z_i is a set consisting of three roommates of unit i. Let $Z = (Z_1, Z_2, \ldots, Z_n)$ be the treatment assignment for all units, and $\mathcal{Z}$ be the set of all possible values of the assignment Z. Let $Y_i(z)$ be the potential outcome of unit i under treatment assignment $z = (z_1, \ldots, z_n)$. This potential outcome depends on treatment assignments of all other units. Let $A_i \in \{1, 2, \ldots, H\}$ be the attribute or type of unit i. In the motivating application, $H = 2$, and $A_i = 1$ if unit i is from Gaokao, and $A_i = 2$ if unit i is from recommendation. Under treatment assignment z, let $R_i(z_i) = \{A_j : j \in z_i\}$ be the set consisting of the attributes of unit i's K peers, and $G_i(z_i) = R_i(z_i) \cup \{A_i\}$ be the set consisting of the attributes of all units in the group that unit i belongs to. We call $R_i(z_i)$ and $G_i(z_i)$ the peer attribute set and group attribute set. Both of them contain unordered but replicable elements. Therefore, $|R_i(z_i)| = K$ and $|G_i(z_i)| = K+1$, where $|\cdot|$ denotes the cardinality of a set. In the motivating application, if unit i is from recommendation and has two roommates from Gaokao and one from recommendation, then $R_i(z_i) = \{1, 1, 2\} \equiv 112$ and $G_i(z_i) = \{1, 1, 2, 2\} \equiv 1122$, where we use 112 and 1122 for notational simplicity. In this case, R_i or G_i has a one-to-one mapping to the number of students from Gaokao within the room of unit i.

Let $I(\cdot)$ be the indicator function. For unit i, $Y_i = Y_i(Z) = \sum_{z \in \mathcal{Z}} I(Z = z) Y_i(z)$ is the observed outcome, and $R_i = R_i(Z_i) = \sum_{z_i} I(Z_i = z_i) R_i(z_i)$ is the observed peer attribute set.

2.2. Group-Level SUTVA and Exclusion-Restriction-Type Assumptions

Without further assumptions, the potential outcome $Y_i(z)$ depends on the treatments of all units. This makes statistical inference intractable. We invoke the following two assumptions to reduce the number of potential outcomes.

Assumption 1. If $z_i = z_i'$, then $Y_i(z) = Y_i(z')$, for any two treatment assignments (z, z') and any unit i.

Assumption 1 states that if a unit's peers do not change, then its potential outcome will not change. This assumption requires no interference between groups but allows for interference within groups. Under Assumption 1, each unit's potential outcomes depend only on its peers in the same group. Therefore, we can write $Y_i(z)$ as $Y_i(z_i)$, a function of the peers of unit i. Assumption 1 is a group-level SUTVA, which is similar to the "partial interference" assumption (Sobel 2006; Hudgens and Halloran 2008).

Assumption 2. If $R_i(z_i) = R_i(z_i')$, then $Y_i(z_i) = Y_i(z_i')$, for any two treatment assignments (z, z') and any unit i.

Assumption 2 states that if the treatment assignment does not affect the attributes of the peers of unit i, then it does not affect the outcome of unit i. Therefore, the potential outcomes of each unit depend only on its peers' attributes instead of its peers' identities. Assumption 2 is similar to "anonymous interaction" (Manski 2013). Assumption 2 implies that the peer attribute set of a unit is the *ultimate treatment* of interest. We are inferring the treatment effects of the peer attribute set. Previous works often invoked Assumption 2, or a slightly weaker form, for inferring peer effects among college roommates. For example, in Sacerdote's (2001) study from Dartmouth College, the ultimate treatment was peers' academic indices created by the admission office, and in Langenskiöld and Rubin's (2008) study from Harvard College, the ultimate treatment was peers' smoking behaviors.

Both Assumptions 1 and 2 are untestable based on the observed data from a single experiment. They are strong identifying assumptions. We will relax them in Section 7.

Under Assumptions 1 and 2, $Y_i(z)$ simplifies to $Y_i(R_i(z_i))$, a function of the peer attribute set of unit i. Recall that $R_i(z_i)$ contains K unordered but replicable elements from $\{1, 2, \ldots, H\}$. Let $\mathcal{R}$ be the set consisting of all possible values of $R_i(z_i)$. Potential outcome of unit i, $Y_i(z)$, simplifies to $Y_i(r)$ for some $r \in \mathcal{R}$. Then the potential outcome is $Y_i(z) = Y_i(R_i(z_i)) = \sum_{r\in\mathcal{R}} I\{R_i(z_i) = r\}Y_i(r)$, and the observed outcome is $Y_i = Y_i(R_i) = \sum_{r\in\mathcal{R}} I(R_i = r)Y_i(r)$. Therefore, we can view the elements in $\mathcal{R}$ as all possible values of the treatment levels, with $|\mathcal{R}| = \binom{K+H-1}{H-1} = \frac{(K+H-1)!}{(H-1)!K!}$ possible values. In our motivating application, $\mathcal{R} = \{r_1, r_2, r_3, r_4\} = \{111, 112, 122, 222\}$ and $|\mathcal{R}| = \frac{(3+2-1)!}{(2-1)!3!} = 4$.

As a side note, motivated by the example of the Chinese university, we consider the case with equal group sizes $K+1$. When groups have different sizes, we need to modify Assumption 2. For example, we can assume that the potential outcomes of a unit depend on the proportions of his/her peers' attributes. The plausibility of this assumption depends on the context of the application, and we leave it to future work.

2.3. Causal Estimands for Peer Effects

For units with attribute $1 \le a \le H$, let $n_{[a]}$ and $w_{[a]} = n_{[a]}/n$ be the number and proportion, and $\bar{Y}_{[a]}(r) = n_{[a]}^{-1}\sum_{i:A_i=a} Y_i(r)$ be the subgroup average potential outcome under treatment r. Let $\bar{Y}(r) = n^{-1}\sum_{i=1}^n Y_i(r)$ be the average potential outcome for all units under treatment r. Therefore, $\bar{Y}(r) = \sum_{a=1}^H w_{[a]}\bar{Y}_{[a]}(r)$ is a weighted average of $\bar{Y}_{[a]}(r)$'s. Comparing treatments $r, r' \in \mathcal{R}$, we define $\tau_i(r, r') = Y_i(r) - Y_i(r')$ as the individual peer effect,

$$\tau_{[a]}(r, r') = n_{[a]}^{-1}\sum_{i:A_i=a} \tau_i(r, r') = \bar{Y}_{[a]}(r) - \bar{Y}_{[a]}(r') \quad (1)$$

as the subgroup average peer effect for units with attribute a, and

$$\tau(r, r') = n^{-1}\sum_{i=1}^n \tau_i(r, r') = \bar{Y}(r) - \bar{Y}(r') = \sum_{a=1}^H w_{[a]}\tau_{[a]}(r, r') \quad (2)$$

as the average peer effect for all units. We are interested in estimating the average peer effects $\tau_{[a]}(r, r')$ and $\tau(r, r')$. They are functions of the fixed potential outcomes and do not depend on the treatment assignment mechanism.

For ease of reading, we summarize the key notation in Table 1.

2.4. Treatment Assignment Mechanism

The treatment assignment mechanism is important for identifying and estimating peer effects. We consider treatment assignment mechanisms satisfying some symmetry conditions. First, units with the same attribute must have the same probability to receive all treatments. Second, pairs of units with the same pair of attributes must have the same probability to receive all pairs of treatments. Formally, we require that the treatment assignment mechanism satisfies the following two conditions.

Assumption 3. For any $r, r' \in \mathcal{R}$,

(a) $\text{pr}(R_i = r) = \text{pr}(R_j = r)$, if $A_i = A_j$;
(b) $\text{pr}(R_i = r, R_j = r') = \text{pr}(R_k = r, R_q = r')$, if $A_i = A_k$ and $A_j = A_q$ for $i \ne j, k \ne q$.

We will give two examples of treatment assignment mechanisms satisfying Assumption 3.

2.4.1. Random Partitioning

Under random partitioning, we randomly assign units to m groups of size $K + 1$, and all possible partitions of units have equal probability. To be more specific, if a treatment assignment z is compatible with a partition of units into m groups of size $K + 1$, then $\text{pr}(Z = z) = m!\{(K + 1)!\}^m/\{m(K + 1)\}!$; otherwise, $\text{pr}(Z = z) = 0$. This formula follows from counting all possible random partitions. To generate a random partition, we can randomly permute $n = m(K+1)$ units and divide them into m groups of equal size $K + 1$ sequentially.

Random partitioning, however, can result in unlucky realizations of the randomization. We may have too few units with attributes and treatments of interest. For illustration, we consider the motivating education example with eight students, five from Gaokao and three from recommendation. Assume that we are interested in $\tau_{[1]}(r_2, r_3)$, the treatment effect of $r_2 = 112$ versus $r_3 = 122$ for students from Gaokao. Under random partitioning, it is possible that no students from Gaokao receives treatment r_2 or r_3. In that case, it is impossible to estimate

Table 1. Notation and explanations.

Notation	Definition	Meaning, properties, or possible values
z_i	peer assignment of unit i	a set of the identity numbers of his/her K peers
A_i	unit i's attribute	$A_i \in \{1, 2, \ldots, H\}$
$n_{[a]}$	number of units with attribute a	$\sum_{a=1}^{H} n_{[a]} = n$
$w_{[a]}$	proportion of units with attribute a	$\sum_{a=1}^{H} w_{[a]} = 1$ and $0 < w_{[a]} < 1 \quad (a = 1, \ldots, H)$
$R_i(z_i)$	unit i's peer attribute set	a set of the attributes of unit i's K peers
$\mathcal{R}$	a set of all possible values of $R_i(z_i)$	$\lvert\mathcal{R}\rvert = \binom{K+H-1}{H-1}$
$G_i(z_i)$	unit i's group attribute set	a set of attributes of all units in unit i's group
$\mathcal{G}$	a set of all possible values of $G_i(z_i)$	$\mathcal{G} = \{g_1, \ldots, g_T\}$ with $T = \lvert\mathcal{G}\rvert = \binom{K+H}{H-1}$
$Y_i(z)$	unit i's potential outcome	under the original treatment $z \in \mathcal{Z}$
$Y_i(r)$	unit i's potential outcome	under the ultimate treatment $r \in \mathcal{R}$

$\tau_{[1]}(r_2, r_3)$ precisely. An example of such a realization is that four students from Gaokao live in one room and the remaining one student from Gaokao and three students from recommendation live in the other room.

2.4.2. Complete Randomization

We propose another treatment assignment mechanism to avoid the drawback of random partitioning. It requires predetermined number of units for each attribute receiving each treatment. We achieve this goal by fixing the numbers of groups. Recall that the group attribute set $G_i(z_i)$ contains $K + 1$ unordered but replicable elements from $\{1, \ldots, H\}$. Consider the same education example with five students from Gaokao and three students from recommendation. We may hope that one room has group attribute set 1112 and thus the other room has group attribute set 1122. This results in three and two students from Gaokao receiving treatments r_2 and r_3, respectively. Therefore, this avoids other assignments with no students from Gaokao receiving these treatments of interest.

We need additional symbols to describe complete randomization. Let $\mathcal{G} = \{g_1, \ldots, g_T\}$ be the set consisting of all possible group attribute sets, with cardinality $T = |\mathcal{G}| = \binom{H+K}{H-1} = \frac{(H+K)!}{(H-1)!(K+1)!}$. In our motivating application, $\mathcal{G} = \{g_1, \ldots, g_5\} = \{1111, 1112, 1122, 1222, 2222\}$ with $T = |\mathcal{G}| = \frac{(2+3)!}{(2-1)!(3+1)!} = 5$. Under treatment assignment z, the number of groups with attribute set $g_t \in \mathcal{G}$ is

$$L_t(z) = (K+1)^{-1} \sum_{i=1}^{n} I\left\{G_i(z_i) = g_t\right\},$$

where the divisor $K + 1$ appears because all $K + 1$ units in the same group must have the same group attribute set. Let $L(z) = (L_1(z), L_2(z), \ldots, L_T(z))$ be the vector of numbers of groups corresponding to group attribute sets $(g_1, \ldots, g_T)$ under assignment z.

Under complete randomization, the assignment z must satisfy $L(z) = l = (l_1, \ldots, l_T)$ for a predetermined constant vector l, and all such assignments must have equal probability. For any $g_t \in \mathcal{G}$, let $g_t(a)$ be the number of elements in set g_t that are equal to a. If z is compatible with a partition of units into m groups and $L(z) = l$, then

$$\text{pr}(Z = z) = \frac{\prod_{t=1}^{T} l_t! \times \prod_{a=1}^{H} \prod_{t=1}^{T} \{g_t(a)!\}^{l_t}}{\prod_{a=1}^{H} n_{[a]}!}; \qquad (3)$$

otherwise, $\text{pr}(Z = z) = 0$. The above formula (3) follows from counting all possible complete randomizations. To generate a complete randomization, we can first randomly partition the $n_{[a]}$ units with attribute a into m groups, where each of the first l_1 groups has $g_1(a)$ units, each of the next l_2 groups has $g_2(a)$ units, ..., each of the last l_T groups has $g_T(a)$ units. The partitions for units with different attributes are mutually independent. Finally, the first l_1 groups will have group attribute set g_1, ..., and the last l_T groups will have group attribute set g_T, satisfying the requirement $L(z) = l$.

We revisit the education example with five students from Gaokao and three students from recommendation. The treatment of complete randomization has predetermined vector $l = (l_1, l_2, l_3, l_4, l_5) = (0, 1, 1, 0, 0)$. Thus, one group has attribute set $g_2 = 1112$ and the other group has attribute set $g_3 = 1122$. We need to randomly assign three students from Gaokao and one student from recommendation to group g_2, and assign the remaining students to group g_3. Equivalently, for the five students from Gaokao, we randomly assign three of them to group g_2 and the remaining two to group g_3; for the three students from recommendation, we randomly assign one of them to group g_2 and the remaining two to group g_3, independently of the group assignments for students from Gaokao.

For $1 \le a \le H$ and $r \in \mathcal{R}$, let $n_{[a]r} = |\{i : A_i = a, R_i = r\}|$ be the number of units with attribute a receiving treatment r. First, the units with attribute a receiving treatment r must have group attribute set $\{a\} \cup r$, which equals g_{t_0} for some $1 \le t_0 \le T$. Second, each group with attribute set g_{t_0} contains $g_{t_0}(a)$ units with attribute a. Third, all of these $g_{t_0}(a)$ units receive the same treatment r. These facts imply that

$$n_{[a]r} = L_{t_0}(z) g_{t_0}(a) = \sum_{t=1}^{T} I(g_t = \{a\} \cup r) \cdot L_t(z) g_t(a) \qquad (4)$$

depends only on the vector $L(z)$. Thus, the $n_{[a]r}$'s are constants under complete randomization. In the previous education example with eight students, consider complete randomization with predetermined vector $l = (0, 1, 1, 0, 0)$. The numbers of units from Gaokao receiving treatments $r_2 = 112$ and $r_3 = 122$ are constants $n_{[1]r_2} = 3$ and $n_{[1]r_3} = 2$. Therefore, complete randomization can guarantee that at least some students from Gaokao receive the treatments of interest.

Moreover, under random partitioning, if we conduct inference conditional on $L(Z)$, then the treatment assignment mechanism becomes complete randomization with $L(z)$ fixed at the observed vector $L(Z)$. Therefore, even under random partitioning, we can still conduct inference under complete randomization if we condition on $L(Z)$.

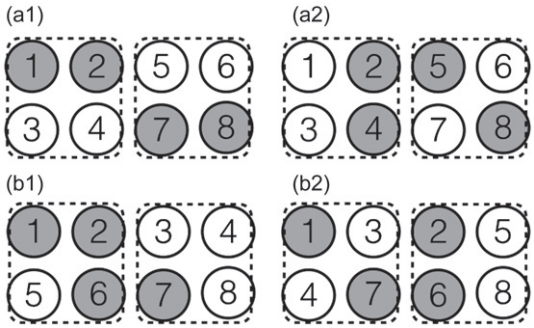


Figure 1. Two types of interference with dashed circles indicating networks. (a) The first type of interference with a fixed network and random external interventions. (a1) and (a2) are two possible realizations, with random external interventions (colors of the units) and a fixed network. (b) The second type of interference with fixed attributes of all units and a random network. (b1) and (b2) are two possible realizations, with fixed attributes (colors of the units) and a random network.

2.5. Connection and Distinction Between Existing Literature and Our Article

We comment on the difference between the majority of the existing literature and our article. We compare two types of interference.

In Figure 1, (a1) and (a2) illustrate the first type. The gray or white color of each unit denotes the external treatment (e.g., receiving vaccine or not). Each unit's outcome depends not only on its own treatment but also on treatments of other units in its circle. Thus, units interfere with each other in the same dashed circle. Importantly, the network structure is fixed.

In Figure 1, (b1) and (b2) illustrate the second type. The gray or white color denotes the units' attributes (e.g., from Gaokao or recommendation in the motivating application). The outcome of each unit depends on the attributes of other units in its circle. Thus, units interfere with each other in the same dashed circle. Unlike the first type, the units' attributes are fixed but the network structure is random.

A main difference between these two types comes from the source of randomness. For the first type, the colors are random and the dashed circles are fixed. For the second type, the colors are fixed and the dashed circles are randomly formed. The recent causal inference literature focused on the first type (Hudgens and Halloran 2008; Aronow 2012; Liu and Hudgens 2014; Athey, Eckles, and Imbens 2018). In this article, we formalize the second type and propose inferential procedures based on the treatment assignment mechanism.

3. Inference for Peer Effects under General Treatment Assignment

3.1. Point Estimators for Peer Effects

Throughout the article, we invoke, unless otherwise stated, Assumptions 1–3. For $1 \le a \le H$ and $r, r' \in \mathcal{R}$, let $\pi_{[a]}(r) = \text{pr}(R_i = r)$ be the probability that a unit i with attribute a receives treatment r. Define

$$\hat{Y}_{[a]}(r) = \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} I(R_i = r)Y_i, \tag{5}$$

$$\hat{\tau}_{[a]}(r,r') = \hat{Y}_{[a]}(r) - \hat{Y}_{[a]}(r'), \quad \hat{\tau}(r,r') = \sum_{a=1}^{H} w_{[a]}\hat{\tau}_{[a]}(r,r'). \tag{6}$$

Proposition 1. For $1 \le a, a' \le H$ and $r, r' \in \mathcal{R}$, the estimators $\hat{Y}_{[a]}(r)$, $\hat{\tau}_{[a]}(r,r')$, and $\hat{\tau}(r,r')$ are unbiased for $\bar{Y}_{[a]}(r)$, $\tau_{[a]}(r,r')$, and $\tau(r,r')$, respectively.

The unbiasedness of $\hat{Y}_{[a]}(r)$ follows from the Horvitz–Thompson-type inverse probability weighting, and the unbiasedness of $\hat{\tau}_{[a]}(r,r')$ and $\hat{\tau}(r,r')$ then follows directly from the linearity of expectation.

3.2. Sampling Variances of the Peer Effect Estimators

For units with attribute $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$, define

$$S^2_{[a]}(r) = (n_{[a]} - 1)^{-1} \sum_{i:A_i=a} \left\{Y_i(r) - \bar{Y}_{[a]}(r)\right\}^2,$$

$$S^2_{[a]}(r\text{-}r') = (n_{[a]} - 1)^{-1} \sum_{i:A_i=a} \left\{\tau_i(r,r') - \tau_{[a]}(r,r')\right\}^2$$

as the finite population variances of the potential outcomes and individual peer effects, and

$$\overline{Y_{[a]}(r)Y_{[a]}(r')} = \{n_{[a]}(n_{[a]} - 1)\}^{-1} \sum_{i\ne j:A_i=A_j=a} Y_i(r)Y_j(r')$$

as the average of the products of the potential outcomes for pairs of units with attribute a.

For $1 \le a, a' \le H$ and $r, r' \in \mathcal{R}$, if $i \ne j$ are two units with attributes a and a', then $\pi_{[a][a']}(r,r') = \text{pr}(R_i = r, R_j = r')$ is the joint treatment assignment probability, and

$$\begin{aligned} d_{[a][a']}(r,r') &= \sqrt{n_{[a]}n_{[a']}}\left\{\frac{\text{pr}(R_i = r, R_j = r')}{\text{pr}(R_i = r)\text{pr}(R_j = r')} - 1\right\} \\ &= \sqrt{n_{[a]}n_{[a']}}\left\{\frac{\pi_{[a][a']}(r,r')}{\pi_{[a]}(r)\pi_{[a']}(r')} - 1\right\} \end{aligned} \tag{7}$$

measures the dependence between the two events $\{R_i = r\}$ and $\{R_j = r'\}$. We further need a few known constants depending only on the treatment assignment mechanism. For $1 \le a, a' \le H$ and $r, r' \in \mathcal{R}$, define

$$c_{[a][a']}(r,r') = \begin{cases} d_{[a][a']}(r,r'), & \text{if } a \ne a', \\ (1 - n^{-1}_{[a]})d_{[a][a]}(r,r') - 1, & \text{if } a = a', r \ne r', \\ (1 - n^{-1}_{[a]})d_{[a][a]}(r,r) + \pi^{-1}_{[a]}(r) - 1, & \text{if } a = a', r = r', \end{cases} \tag{8}$$

and

$$b_{[a]}(r) = (1 - n^{-1}_{[a]})\left\{c_{[a][a]}(r,r) - d_{[a][a]}(r,r)\right\} + 1. \tag{9}$$

These constants are useful for expressing the sampling variances of the estimators in (6).

Theorem 1. Under Assumptions 1–3, for treatments $r \ne r' \in \mathcal{R}$, the sampling variance of the subgroup average peer effect estimator is

$$\text{var}\left\{\hat{\tau}_{[a]}(r,r')\right\} = n^{-1}_{[a]}\left\{b_{[a]}(r)S^2_{[a]}(r) + b_{[a]}(r')S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r')\right\}$$

$$+ n_{[a]}^{-1}\Big\{c_{[a][a]}(r,r)\bar{Y}^2_{[a]}(r) + c_{[a][a]}(r',r')\bar{Y}^2_{[a]}(r') - 2c_{[a][a]}(r,r')\overline{Y_{[a]}(r)Y_{[a]}(r')}\Big\}, \quad (10)$$

and the sampling variance of the average peer effect estimator is

$$\begin{aligned}
&\text{var}\left\{\hat{\tau}(r,r')\right\} \\
&= n^{-1}\sum_{a=1}^{H} w_{[a]}\left\{b_{[a]}(r)S^2_{[a]}(r) + b_{[a]}(r')S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r')\right\} \\
&\quad + n^{-1}\sum_{a=1}^{H} w_{[a]}\Big\{c_{[a][a]}(r,r)\bar{Y}^2_{[a]}(r) + c_{[a][a]}(r',r')\bar{Y}^2_{[a]}(r') - 2c_{[a][a]}(r,r')\overline{Y_{[a]}(r)Y_{[a]}(r')}\Big\} \\
&\quad + n^{-1}\sum_{a=1}^{H}\sum_{a'\neq a}(w_{[a]}w_{[a']})^{1/2}\big\{c_{[a][a']}(r,r)\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r) + c_{[a][a']}(r',r')\bar{Y}_{[a]}(r')\bar{Y}_{[a']}(r') \\
&\quad - c_{[a][a']}(r,r')\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r') - c_{[a][a']}(r',r)\bar{Y}_{[a]}(r')\bar{Y}_{[a']}(r)\big\}. \quad (11)
\end{aligned}$$

From Theorem 1, the sampling variances of the peer effect estimators depend on the finite population variances of potential outcomes and individual peer effects, the products of two subgroup average potential outcomes, and the product averages $\overline{Y_{[a]}(r)Y_{[a]}(r')}$'s. In contrast to $\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')$, the average $\overline{Y_{[a]}(r)Y_{[a]}(r')}$ excludes the product of two potential outcomes of the same unit. Note that we cannot unbiasedly estimate quantities involving $Y_i(r)Y_i(r')$ in general because we cannot jointly observe the potential outcomes, $Y_i(r)$ and $Y_i(r')$, for any unit i and any treatments $r \neq r'$.

Moreover, the sampling variance of $\hat{\tau}(r,r')$ is a weighted summation of the sampling variances of the $\hat{\tau}_{[a]}(r,r')$'s, corresponding to the first two terms in (11), and the sampling covariances between $\hat{\tau}_{[a]}(r,r')$ and $\hat{\tau}_{[a']}(r,r')$, corresponding to the last double summation in (11).

3.3. Estimating the Sampling Variances

From Theorem 1, to estimate the sampling variances, we need to estimate the population quantities in (10) and (11). For $1 \leq a \leq H$, define

$$s^2_{[a]}(r) = \frac{n_{[a]}\pi^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r)} \times \left\{\frac{n_{[a]} + c_{[a][a]}(r,r)}{n^2_{[a]}\pi_{[a]}(r)}\sum_{i:A_i=a} I(R_i = r)Y_i^2 - \hat{Y}^2_{[a]}(r)\right\}. \quad (12)$$

The following theorem is the basis for constructing variance estimators.

Theorem 2. Under Assumptions 1–3, for $1 \leq a, a' \leq H$ and $r, r' \in \mathcal{R}$,

$$S^2_{[a]}(r) = E\left\{s^2_{[a]}(r)\right\},$$

$$\bar{Y}^2_{[a]}(r) = E\left[\frac{n_{[a]}\hat{Y}^2_{[a]}(r) - \{b_{[a]}(r)-1\}s^2_{[a]}(r)}{n_{[a]} + c_{[a][a]}(r,r)}\right],$$

$$\overline{Y_{[a]}(r)Y_{[a]}(r')} = E\left\{\frac{n_{[a]}}{n_{[a]}-1}\frac{\pi_{[a]}(r)\pi_{[a]}(r')}{\pi_{[a][a]}(r,r')}\hat{Y}_{[a]}(r)\hat{Y}_{[a]}(r')\right\}, \quad \text{if } r \neq r',$$

$$\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r') = E\left\{\frac{\pi_{[a]}(r)\pi_{[a']}(r')}{\pi_{[a][a']}(r,r')}\hat{Y}_{[a]}(r)\hat{Y}_{[a']}(r')\right\}, \quad \text{if } a \neq a'.$$

The estimators in Theorem 2 correspond to the sample analogs of these finite population quantities, with carefully chosen coefficients to ensure unbiasedness. Theorem 2 guarantees that we have unbiased estimators for all terms in $\text{var}\{\hat{\tau}_{[a]}(r,r')\}$ and $\text{var}\{\hat{\tau}(r,r')\}$ except the variance of the individual peer effects $S^2_{[a]}(r\text{-}r')$. We cannot unbiasedly estimate $S^2_{[a]}(r\text{-}r')$ from the observed data. This is analogous to other finite population causal inference (Neyman 1923). Because the coefficients of $S^2_{[a]}(r\text{-}r')$ in the variance formulas (10) and (11) are both negative, we can ignore the terms involving $S^2_{[a]}(r\text{-}r')$ and conservatively estimate the sampling variances by simply plugging in the estimators in Theorem 2. Note that $S^2_{[a]}(r\text{-}r') = 0$ holds under *additivity* defined below.

Definition 1. The individual peer effects for units with attribute a are additive if and only if $\tau_i(r,r') = Y_i(r) - Y_i(r')$ is constant for each unit i with attribute a, or, equivalently, $S^2_{[a]}(r\text{-}r') = 0$.

Therefore, the final estimator for $\text{var}\{\hat{\tau}_{[a]}(r,r')\}$ is unbiased under additivity for a, and the final estimator for $\text{var}\{\hat{\tau}(r,r')\}$ is unbiased under additivity for all $1 \leq a \leq H$.

4. Inference for Peer Effects under Complete Randomization

Under random partitioning, the formulas of $\pi_{[a]}(r)$, $\pi_{[a][a']}(r,r')$, $d_{[a][a']}(r,r')$, $b_{[a]}(r)$, and $c_{[a][a']}(r,r')$ are complicated, and so are the sampling variances of peer effect estimators. We relegate them to the supplementary material. Fortunately, they have much simpler forms under complete randomization. In this section, we will focus on the inference under complete randomization.

4.1. Treatment Assignment under Complete Randomization

The randomness in the peer effect estimators comes solely from the treatment assignments for all units, $(R_1, \ldots, R_n)$. Therefore, we need to first characterize the distribution of the treatments under complete randomization. Intuitively, the symmetry of complete randomization suggests that $(R_1, \ldots, R_n)$ has the same distribution as the treatment of a stratified randomized experiment. The following proposition states this equivalence formally.

Proposition 2. Under Assumptions 1 and 2, the complete randomization defined in Section 2.4.2 induces a stratified random-

ized experiment, in the sense that (1) for each $1 \le a \le H$, in the stratum consisting of $n_{[a]}$ units with attribute a, $n_{[a]r}$ units receive treatment r for any $r \in \mathcal{R}$, and any realization of treatments for these $n_{[a]}$ units has the same probability; and (2) the treatments of units are independent across strata.

Proposition 2 follows from the numerical implementation of the complete randomization described in Section 2.4.2. It implies the formulas of $\pi_{[a]}(r)$, $\pi_{[a][a']}(r,r')$, $d_{[a][a']}(r,r')$, $b_{[a]}(r)$, and $c_{[a][a']}(r,r')$. We give a formal proof in the supplementary material. The group assignment for units with the same attribute a induces a completely randomized experiment, with $n_{[a]r}$ units receiving treatment r. Moreover, the group assignments for units with different attributes are mutually independent.

4.2. Point Estimators for Peer Effects

Proposition 2 characterizes the treatment assignment of complete randomization, which allows us to express the peer effect estimators in simpler forms.

Corollary 1. Under Assumptions 1 and 2, and under the complete randomization defined in Section 2.4.2, for $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$,

$$\hat{Y}_{[a]}(r) = n_{[a]r}^{-1} \sum_{i:A_i=a,R_i=r} Y_i, \quad \hat{\tau}_{[a]}(r,r') = \hat{Y}_{[a]}(r) - \hat{Y}_{[a]}(r'),$$
$$\hat{\tau}(r,r') = \sum_{a=1}^{H} w_{[a]} \hat{\tau}_{[a]}(r,r'). \tag{13}$$

Therefore, under complete randomization, the unbiased estimator of the subgroup average peer effect, $\hat{\tau}_{[a]}(r,r')$, is the observed difference in outcome means under treatments r and r' for units with attribute a.

4.3. Sampling Variances of the Peer Effect Estimators

The sampling variances also have simpler forms under complete randomization.

Corollary 2. Under Assumptions 1 and 2, and under the complete randomization defined in Section 2.4.2, for $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$,

$$\text{var}\left\{\hat{\tau}_{[a]}(r,r')\right\} = \frac{S^2_{[a]}(r)}{n_{[a]r}} + \frac{S^2_{[a]}(r')}{n_{[a]r'}} - \frac{S^2_{[a]}(r\text{-}r')}{n_{[a]}},$$
$$\text{var}\left\{\hat{\tau}(r,r')\right\} = \sum_{a=1}^{H} w_{[a]}^2 \left\{\frac{S^2_{[a]}(r)}{n_{[a]r}} + \frac{S^2_{[a]}(r')}{n_{[a]r'}} - \frac{S^2_{[a]}(r\text{-}r')}{n_{[a]}}\right\}.$$

From Corollary 2, the variance formula of the subgroup average peer effect estimator under complete randomization is the same as that for classical completely randomized experiments with multiple treatments (Neyman 1923). This follows from the equivalence relationship in Proposition 2. Corollary 2 also implies that $\text{var}\{\hat{\tau}(r,r')\} \equiv \text{var}\{\sum_{a=1}^{H} w_{[a]}\hat{\tau}_{[a]}(r,r')\} = \sum_{a=1}^{H} w_{[a]}^2 \text{var}\{\hat{\tau}_{[a]}(r,r')\}$. This follows from the mutual independence of $\{\hat{\tau}_{[a]}(r,r') : 1 \le a \le H\}$ in an experiment stratified on attributes.

From Corollary 2, the $n_{[a]r}$'s are the effective sample sizes. On the one hand, this is intuitive because they are the sample sizes of the stratified experiment described in Proposition 2. On the other hand, this is counterintuitive because units in the same group have correlated observed outcomes. However, this correlation does not diminish the effective sample sizes in contrast to the correlation in standard group-randomized experiments. Units in the same group could potentially be in a different group under a different realization of the treatment assignment. The probability that two given units are in the same group decreases as n increases, and so does the correlation between their observed outcomes. To repeat, under complete randomization, we analyze the data as if they are from a stratified experiment, not from a group-randomized experiment.

4.4. Estimating the Sampling Variances

From Proposition 2, $\hat{Y}_{[a]}(r) = \sum_{i:A_i=a,R_i=r} Y_i/n_{[a]r}$ reduces to the sample mean, and

$$s^2_{[a]}(r) = \frac{n_{[a]r}}{n_{[a]r}-1}\left\{\frac{1}{n_{[a]r}} \sum_{i:A_i=a,R_i=r} Y_i^2 - \hat{Y}^2_{[a]}(r)\right\}$$
$$= (n_{[a]r}-1)^{-1} \sum_{i:A_i=a,R_i=r} \left\{Y_i - \hat{Y}_{[a]}(r)\right\}^2 \tag{14}$$

reduces to the sample variance of the observed outcomes for units with attribute a receiving treatment r. Formula (14), coupled with Corollary 2, simplifies the variance estimators under complete randomization, which coincide with Neyman's (1923) conservative variance estimators under classical completely randomized experiments with multiple treatments.

Corollary 3. Under Assumptions 1 and 2, and under the complete randomization defined in Section 2.4.2, for $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$, the variance estimators become

$$\hat{V}_{[a]}(r,r') = \frac{s^2_{[a]}(r)}{n_{[a]r}} + \frac{s^2_{[a]}(r')}{n_{[a]r'}},$$
$$\hat{V}(r,r') = \sum_{a=1}^{H} w_{[a]}^2 \left\{\frac{s^2_{[a]}(r)}{n_{[a]r}} + \frac{s^2_{[a]}(r')}{n_{[a]r'}}\right\}. \tag{15}$$

Moreover, $E\{\hat{V}_{[a]}(r,r')\} - \text{var}\{\hat{\tau}_{[a]}(r,r')\} = n_{[a]}^{-1} S^2_{[a]}(r\text{-}r') \ge 0$, which becomes zero under additivity for a, and $E\{\hat{V}(r,r')\} - \text{var}\{\hat{\tau}(r,r')\} = n^{-1}\sum_{a=1}^{H} w_{[a]} S^2_{[a]}(r\text{-}r') \ge 0$, which becomes zero under additivity for all $1 \le a \le H$.

4.5. Asymptotic Distributions and Confidence Intervals for Peer Effects

The asymptotic analysis embeds the n units into a sequence of finite populations with increasing sizes. See Li and Ding (2017) for a review of finite population asymptotics in causal inference. Under complete randomization, if some regularity conditions hold, then $\hat{\tau}_{[a]}(r,r')$ is asymptotically Normal. We can then construct a $1-\alpha$ Wald-type confidence interval for $\tau_{[a]}(r,r')$: $\hat{\tau}_{[a]}(r,r') \pm q_{1-\alpha/2}\hat{V}^{1/2}_{[a]}(r,r')$, with $q_{1-\alpha/2}$ being the

$(1-\alpha/2)$th quantile of $\mathcal{N}(0,1)$. Because the variance estimator $\hat{V}_{[a]}(r,r')$ in (15) overestimates the true sampling variance on average, the confidence interval is asymptotically conservative, with the limit of coverage probability larger than or equal to the nominal level. Analogously, we can construct asymptotically conservative confidence intervals for $\tau(r,r')$. We formally state the regularity condition as follows.

Condition 1. For any $1 \le a \le H, r \ne r' \in \mathcal{R}$, as $n \to \infty$,

(i) the proportions, $w_{[a]}$ and $n_{[a]r}/n_{[a]}$, have positive limits,
(ii) the finite population variances of potential outcomes and individual peer effects, $S^2_{[a]}(r)$ and $S^2_{[a]}(r\text{-}r')$, have limits, and at least one of the limits of $\{S^2_{[a]}(r) : r \in \mathcal{R}\}$ are nonzero,
(iii) $\max_{i:A_i=a} |Y_i(r) - \bar{Y}_{[a]}(r)|^2/n_{[a]} \to 0$.

Conditions (i) and (ii) are natural in most applications. In our motivating application, GPA is bounded within $[0,4]$, and therefore condition (iii) holds automatically (Li and Ding 2017). We summarize the asymptotic results below.

Theorem 3. Under Assumptions 1 and 2, and under the complete randomization defined in Section 2.4.2, if Condition 1 holds, then, for any $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$,

(a) $\hat{\tau}_{[a]}(r,r')$ and $\hat{\tau}(r,r')$ are asymptotically Normal,
(b) the Wald-type confidence intervals for $\tau_{[a]}(r,r')$ and $\tau(r,r')$ are asymptotically conservative, unless the peer effects are additive for units with the same attribute.

4.6. Randomization-Based and Regression-Based Analyses

Theorem 3 is purely randomization based without any modeling assumptions of the outcomes. Regression-based analysis is also popular in practice. Now we make a connection between them. Suppose we fit a linear model for the observed outcomes

$$Y_i = \mu + \alpha_{[A_i]} + \beta_{R_i} + \lambda_{[A_i]R_i} + \varepsilon_i, \tag{16}$$

where μ is the intercept, $\alpha_{[a]}$ represents the main "effect" of attribute a, β_r represents the main effect of treatment r, and $\lambda_{[a]r}$ represents the interaction between attribute a and treatment r. The traditional linear regression assumes that the error terms follow independent zero-mean (Normal) distributions and generate the randomness of the observed outcomes.

Under model (16), we need some constraints to avoid over-parameterization: $\sum_{a=1}^H \alpha_{[a]} = 0$, $\sum_{r\in\mathcal{R}} \beta_r = 0$, $\sum_{a=1}^H \lambda_{[a]r} = 0$, and $\sum_{r\in\mathcal{R}} \lambda_{[a]r} = 0$, for $1 \le a \le H$ and $r \in \mathcal{R}$. Let $\mu_{[a]r} = \mu + \alpha_{[a]} + \beta_r + \lambda_{[a]r}$. Then we can interpret $\mu_{[a]r} - \mu_{[a]r'}$ as the subgroup average peer effect of treatment r versus r' for units with attribute a. The least-squares estimators of the coefficients are

$$\hat{\mu} = \frac{1}{H|\mathcal{R}|}\sum_{a=1}^H\sum_{r\in\mathcal{R}} \hat{Y}_{[a]}(r), \quad \hat{\alpha}_{[a]} = \frac{1}{|\mathcal{R}|}\sum_{r\in\mathcal{R}} \hat{Y}_{[a]}(r) - \hat{\mu},$$

$$\hat{\beta}_r = \frac{1}{H}\sum_{a=1}^H \hat{Y}_{[a]}(r) - \hat{\mu}, \qquad \hat{\lambda}_{[a]r} = \hat{Y}_{[a]}(r) - (\hat{\mu} + \hat{\alpha}_{[a]} + \hat{\beta}_r),$$

$$(1 \le a \le H, r \in \mathcal{R}).$$

Then we have the following proposition.

Proposition 3. Under the linear model (16), for any $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$, the least-squares estimator for the subgroup average peer effect is

$$\begin{aligned}\hat{\mu}_{[a]r} - \hat{\mu}_{[a]r'} &= (\hat{\mu} + \hat{\alpha}_{[a]} + \hat{\beta}_r + \hat{\lambda}_{[a]r}) \\ &\quad - (\hat{\mu} + \hat{\alpha}_{[a]} + \hat{\beta}_{r'} + \hat{\lambda}_{[a]r'}) = \hat{Y}_{[a]}(r) - \hat{Y}_{[a]}(r') \\ &= \hat{\tau}_{[a]}(r,r'),\end{aligned}$$

with the Huber–White variance estimator

$$\begin{aligned}\hat{V}_{[a],\mathrm{HW}}(r,r') &= \frac{n_{[a]r}-1}{n_{[a]r}}\frac{s^2_{[a]}(r)}{n_{[a]r}} + \frac{n_{[a]r'}-1}{n_{[a]r'}}\frac{s^2_{[a]}(r')}{n_{[a]r'}} \\ &\approx \frac{s^2_{[a]}(r)}{n_{[a]r}} + \frac{s^2_{[a]}(r')}{n_{[a]r'}} = \hat{V}_{[a]}(r,r').\end{aligned}$$

The linear outcome model (16) includes the *interaction* between the unit's attribute A_i and the treatment received R_i. Under (16), both the point estimator and Huber–White variance estimator for the subgroup average peer effect are (nearly) identical to the randomization-based ones under complete randomization. Complete randomization justifies this regression-based analysis for peer effects. This result extends Lin (2013) for covariate adjustment in classical completely randomized experiments. However, such equivalence generally does not hold if the treatment assignment is not complete randomization, nor if we use the conventional variance estimator in linear models assuming homoscedasticity of the error terms.

Related to the discussion of effective sample sizes after Proposition 2, we do *not* need to use cluster-robust standard errors even though some units are in the same group or cluster. Our inference depends solely on the random assignment of peers in contrast to model-based inferences (e.g., Carrell, Sacerdote, and West 2013). In our setting, randomization does *not* justify cluster-robust standard errors. Our view is similar to Abadie et al. (2017) in a different context.

Many econometric analyses of peer effects did not include the interaction term (e.g., Sacerdote 2001; Carrell, Sacerdote, and West 2013). Complete randomization does *not* justify them in the presence of treatment effect heterogeneity. Sometimes, peers' outcomes also enter the right-hand side of the regression in (16). It is then more difficult to interpret their least-squares coefficients as causal effects estimators (Manski 1993; Angrist 2014).

4.7. Asymptotic Distributions and Confidence Sets for Multiple Peer Effects

Below we study the joint asymptotic sampling distribution of multiple average peer effect estimators. It is useful for constructing confidence sets and testing significance of multiple average peer effects simultaneously. For mathematical convenience, we center the potential outcomes:

$$\theta_i(r) = Y_i(r) - |\mathcal{R}|^{-1}\sum_{r'\in\mathcal{R}} Y_i(r'),$$

$$\theta_{[a]}(r) = n_{[a]}^{-1}\sum_{i:A_i=a}\theta_i(r) = \bar{Y}_{[a]}(r) - |\mathcal{R}|^{-1}\sum_{r'\in\mathcal{R}}\bar{Y}_{[a]}(r'),$$

$$\theta_i(\mathcal{R}) = (\theta_i(r_1), \ldots, \theta_i(r_{|\mathcal{R}|}))^\top,$$
$$\theta_{[a]}(\mathcal{R}) = (\theta_{[a]}(r_1), \ldots, \theta_{[a]}(r_{|\mathcal{R}|}))^\top.$$

Let $\hat{\theta}_{[a]}(r) = \hat{Y}_{[a]}(r) - |\mathcal{R}|^{-1}\sum_{r'\in\mathcal{R}} \hat{Y}_{[a]}(r')$ be the centered subgroup average potential outcome estimator, vectorized as $\hat{\theta}_{[a]}(\mathcal{R}) = (\hat{\theta}_{[a]}(r_1), \ldots, \hat{\theta}_{[a]}(r_{|\mathcal{R}|}))^\top$. For any $r, r' \in \mathcal{R}$, the individual peer effect $\tau_i(r, r')$, the subgroup average peer effect $\tau_{[a]}(r, r')$, and the subgroup average peer effect estimator $\hat{\tau}_{[a]}(r, r')$ are the same linear transformations of $\theta_i(\mathcal{R}), \theta_{[a]}(\mathcal{R})$, and $\hat{\theta}_{[a]}(\mathcal{R})$, respectively. Therefore, it suffices to study the joint asymptotic sampling distribution of the $\hat{\theta}_{[a]}(\mathcal{R})$'s for all a, and construct confidence sets for $\theta_{[a]}(\mathcal{R})$'s and their linear transformations.

Define $\Gamma = I_{|\mathcal{R}|} - |\mathcal{R}|^{-1} 1_{|\mathcal{R}|} 1_{|\mathcal{R}|}^\top$ as an $|\mathcal{R}| \times |\mathcal{R}|$ projection matrix orthogonal to $1_{|\mathcal{R}|}$. The theorem below summarizes the results for the joint inference of multiple peer effects.

Theorem 4. Under Assumptions 1 and 2, the complete randomization defined in Section 2.4.2, and Condition 1, (a) $\hat{\theta}_{[1]}(\mathcal{R}), \ldots, \hat{\theta}_{[H]}(\mathcal{R})$ are mutually independent; (b) $\hat{\theta}_{[a]}(\mathcal{R})$ is unbiased for $\theta_{[a]}(\mathcal{R})$ with sampling covariance $\mathrm{cov}\{\hat{\theta}_{[a]}(\mathcal{R})\}$ as follows:

$$\Gamma\, \mathrm{diag}\left\{n_{[a]r_1}^{-1} S_{[a]}^2(r_1), \ldots, n_{[a]r_{|\mathcal{R}|}}^{-1} S_{[a]}^2(r_{|\mathcal{R}|})\right\}\Gamma - \frac{1}{n_{[a]}(n_{[a]}-1)} \sum_{i:A_i=a} \left\{\theta_i(\mathcal{R}) - \theta_{[a]}(\mathcal{R})\right\} \times \left\{\theta_i(\mathcal{R}) - \theta_{[a]}(\mathcal{R})\right\}^\top;$$

(c) $\hat{\theta}_{[a]}(\mathcal{R}) - \theta_{[a]}(\mathcal{R})$ is asymptotically Normal with mean 0 and covariance $\mathrm{cov}\{\hat{\theta}_{[a]}(\mathcal{R})\}$; (d) the covariance estimator

$$\widehat{\mathrm{cov}}\{\hat{\theta}_{[a]}(\mathcal{R})\} = \Gamma\, \mathrm{diag}\left\{n_{[a]r_1}^{-1} s_{[a]}^2(r_1), \ldots, n_{[a]r_{|\mathcal{R}|}}^{-1} s_{[a]}^2(r_{|\mathcal{R}|})\right\}\Gamma \tag{17}$$

is conservative in expectation, unless the peer effects are additive for units with attribute a.

From Theorem 4, we can then obtain the Wald-type asymptotic conservative confidence sets for $(\theta_{[1]}(\mathcal{R})^\top, \ldots, \theta_{[H]}(\mathcal{R})^\top)^\top$ and their linear transformations, including multiple average or subgroup average peer effects as special cases.

5. Optimal Treatment Assignment Mechanism

5.1. Point Estimator for the Optimal Treatment Assignment Mechanism

The results in previous sections are useful for decision-making. We can use them to find the optimal treatment assignment mechanism for a new population of size $n' = m'(K+1)$. We need to assume that the new population is similar to the one in our data in some way. Otherwise, we cannot draw any conclusions in general. For instance, we assume that the subgroup average potential outcomes in the new population are linear transformations of those in our data, that is, $\bar{Y}'_{[a]}(r) = c\bar{Y}_{[a]}(r) + \xi_{[a]}$ for some $c > 0$ and $\xi_{[a]}$, for all $1 \le a \le H$ and $r \in \mathcal{R}$. In our motivating application, the new population usually consists of the students coming next year. The scale parameter c and shift parameter $\xi_{[a]}$ can explain the proportional change and the absolute change of average GPAs across different years. These changes are possibly due to the difference in qualities of students and difficulties of exams across years.

We use complete randomization with $L'(z)$ fixed at some vector l' for the new population. Our goal is to find $l' = (l'_1, \ldots, l'_T)$ to maximize the expected total outcome. We are looking for the optimal l' of complete randomization, and the final assignment Z' is still random. For any $1 \le a \le H$ and $r \in \mathcal{R}$, let $n'_{[a]r}$ be the number of units with attribute a receiving treatment r in the new population under complete randomization with $L'(z)$ fixed at l'. Proposition 2 and (4) have the following useful implications. First, $n'_{[a]r}$ is a deterministic function of l'. Second, within each stratum consisting of $n'_{[a]}$ units with attribute a, we randomly assign $n'_{[a]r}$ units to treatment r for any $r \in \mathcal{R}$. Third, the treatments for units in different strata are mutually independent. Based on these, the expected total outcome under complete randomization with $L'(z) = l'$ is

$$E\left(\sum_{i=1}^{n'} Y'_i\right) = \sum_{a=1}^{H} E\left(\sum_{i:A_i=a} Y'_i\right) = \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}'_{[a]}(r) = c\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}_{[a]}(r) + \sum_{a=1}^{H} n'_{[a]}\xi_{[a]}, \tag{18}$$

where the last equality follows from $\bar{Y}'_{[a]}(r) = c\bar{Y}_{[a]}(r) + \xi_{[a]}$ and $\sum_{r\in\mathcal{R}} n'_{[a]r} = n'_{[a]}$. Although the expected total outcome (18) of the new population depends on the unknown constants $c > 0$ and $\xi_{[a]}$'s, the maximizer l'_{opt} for this expected total outcome is the same as that for $\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}_{[a]}(r)$. Moreover, we can unbiasedly estimate $\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}_{[a]}(r)$ by replacing $\bar{Y}_{[a]}(r)$ with the corresponding unbiased estimator $\hat{Y}_{[a]}(r)$, and then use $\hat{l}'_{\mathrm{opt}}$ that maximizes the unbiased estimator $\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\hat{Y}_{[a]}(r)$ as an estimator for l'_{opt}. From (4), the objective function reduces to

$$\begin{aligned}\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\hat{Y}_{[a]}(r) &= \sum_{a=1}^{H}\sum_{r\in\mathcal{R}}\left\{\sum_{t=1}^{T} I(g_t = \{a\}\cup r) l'_t g_t(a)\right\}\hat{Y}_{[a]}(r)\\ &= \sum_{t=1}^{T}\left\{\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} I(g_t = \{a\}\cup r) g_t(a)\hat{Y}_{[a]}(r)\right\} l'_t,\end{aligned} \tag{19}$$

which is a linear function of $l' = (l'_1, \ldots, l'_T)$. In (19), the coefficient of l'_t is the estimated total outcome of $K+1$ units in the group with group attribute g_t. The constraints on l' include that all the l'_t's are nonnegative integers, and the number of units with attribute a implied by l' is fixed:

$$\begin{cases}\sum_{t=1}^{T} g_t(a) l'_t = n'_{[a]}, & (a = 1, \ldots, H),\\ l'_t \ge 0, \quad l'_t \text{ is an integer}, & (t = 1, \ldots, T).\end{cases} \tag{20}$$

Both the objective function (19) and the constraints (20) are linear in l'. Therefore, finding the maximizer $\hat{l}'$ is a linear integer programming problem. When the sample size is not too large,

we can enumerate all possible values of l' to obtain the maximizer.

Bhattacharya (2009) discussed the optimal peer assignment in a super population scenario where each unit has only one peer ($K = 1$). In that case, the optimization problem becomes a linear programming problem without the integer constraint.

5.2. Inference for the Optimal Assignment Mechanism

Section 5.1 gives a point estimator of the optimal treatment assignment mechanism. The uncertainty of the point estimator comes from the uncertainty of the $\hat{Y}_{[a]}(r)$'s. Below we construct confidence sets for the optimal l'_{opt} of complete randomization. Note that $\sum_{r\in\mathcal{R}} n'_{[a]r} = n'_{[a]}$ is fixed for all a. By the definitions of $\theta_{[a]}(r)$ and $\hat{\theta}_{[a]}(r)$, the true and estimated optimal treatment assignment mechanisms satisfy

$$\begin{aligned} l'_{\text{opt}} &= \arg\max_{l'\in\mathcal{L}'} \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}_{[a]}(r) \\ &= \arg\max_{l'\in\mathcal{L}'} \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\theta_{[a]}(r), \qquad (21) \\ \hat{l}'_{\text{opt}} &= \arg\max_{l'\in\mathcal{L}'} \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\hat{Y}_{[a]}(r) \\ &= \arg\max_{l'\in\mathcal{L}'} \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\hat{\theta}_{[a]}(r), \end{aligned}$$

where $\mathcal{L}'$ denotes the set of all possible l' satisfying the constraint (20). Here, we represent l'_{opt} using the centered subgroup average potential outcomes, because the $\theta_{[a]}(r)$'s have simpler asymptotically conservative confidence sets, as shown in Theorem 4. For any $\alpha \in (0,1)$, let $\mathcal{C}_{[a]}(\alpha)$ be the $1-\alpha$ Wald-type asymptotic conservative confidence set for $\theta_{[a]}(\mathcal{R})$. Then a $1-\alpha$ asymptotic conservative confidence set for l'_{opt} is

$$\left\{ \arg\max_{l'\in\mathcal{L}'} \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{\theta}_{[a]}(r) : \bar{\theta}_{[1]}(\mathcal{R}) \in \mathcal{C}_{[1]}(\bar{\alpha}), \ldots, \bar{\theta}_{[H]}(\mathcal{R}) \in \mathcal{C}_{[H]}(\bar{\alpha}), \bar{\alpha} = 1-(1-\alpha)^{1/H} \right\}. \qquad (22)$$

The confidence set in (22) involves solving infinite linear integer programming problems, which are computationally intensive. More importantly, the interpretation of the confidence set in (22) seems unnatural for making decisions in the future because the "confidence" statement is a property over repeated sampling of the previous experiment. Ideally, we need to make future decisions conditioning on the observed data rather than averaging over them. Below we use the "fiducial distribution" (Fisher 1935; Dasgupta, Pillai, and Rubin 2015) of l'_{opt}.

We start with the asymptotic sampling distribution in Theorem 4, and then swap the roles of the estimators and estimands. Let $\tilde{\theta}_{[a]}(\mathcal{R}) \equiv (\tilde{\theta}_{[a]}(r_1), \ldots, \tilde{\theta}_{[a]}(r_{|\mathcal{R}|}))$ be a multivariate Normal distribution with mean $\hat{\theta}_{[a]}(\mathcal{R})$ and covariance $\widehat{\text{cov}}\{\hat{\theta}_{[a]}(\mathcal{R})\}$ in (17), independently for $1 \le a \le H$. The fiducial distribution of l'_{opt} is the distribution of $\tilde{l}'_{\text{opt}} \equiv \arg\max_{l'\in\mathcal{L}'} \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\tilde{\theta}_{[a]}(r)$, with $\theta_{[a]}(r)$ in (21) replaced by the *random variable* $\tilde{\theta}_{[a]}(r)$. When there exist multiple maximizers for $\tilde{l}'_{\text{opt}}$, we randomly choose one of them with equal probability. Computationally, to simulate from $\tilde{l}'_{\text{opt}}$, we can first simulate the $\tilde{\theta}_{[a]}(\mathcal{R})$'s independently from Normal distributions and then calculate $\tilde{l}'_{\text{opt}}$. Compared to confidence sets, the "fiducial distribution" not only acts as a computational compromise but also has a natural Bayesian interpretation. We can view $\tilde{\theta}_{[a]}(\mathcal{R})$ as the Bayesian posterior distribution of $\theta_{[a]}(\mathcal{R})$ based on the sampling distribution in Theorem 4 under a flat prior. Consequently, $\tilde{l}'_{\text{opt}}$ is the Bayesian posterior distribution of l'_{opt}. Rigorously, this is not a full Bayesian procedure but only a limited information Bayesian procedure (Kwan 1999; Sims 2006). It uses "limited information" from the asymptotic randomization distribution, but it does not impose a full outcome model for all units. Therefore, this "fiducial distribution" enjoys not only the robustness of randomization inference without outcome modeling but also the interpretability of Bayesian inference for decision-making.

6. Application to Roommate Assignment in a Top Chinese University

6.1. Overview of the Data

The dataset consists of college students graduating in a recent year from 25 departments of a top Chinese university. The university assigns dorms to departments, and the departments then assign students to dorms. The roommate assignment is close to random partitioning for students of the same department, gender, and graduating year.

Among these students, 73.9% of them were from Gaokao, 21.7% were from recommendation, and 4.4% were from other ways (omitted in our analysis). For example, 43.2% students are from recommendation in the mathematics department, 45.2% in the physics department, 52.5% in the chemistry department, 22.4% in the biology department, and 28.4% in the informatics department. The outcome is the freshman year GPA. The average freshman GPAs are 3.26 and 3.37 for students from Gaokao and recommendation, respectively. On average, students from recommendation do better than students from Gaokao during the freshman year. We first want to understand the effect of roommate types on students' academic performance. We then need to design an optimal roommate assignment.

6.2. Point and Interval Estimators for Peer Effects

The student room assignment is conditional on department, gender, and the graduating year. We focus on the male students from the Departments of Informatics and Physics separately. These two departments have larger sample sizes. Moreover, there is a close connection between the training in high school Olympiads and the freshman introductory courses in these two departments. The informatics department has 104 students from Gaokao and 52 students from recommendation. The physics department has 49 students from Gaokao and 43 students from recommendation. If we conduct inference conditioning on the numbers of groups with different group attribute

Table 2. Estimated peer effects with $\mathcal{R} = \{r_1, r_2, r_3, r_4\} = \{111, 112, 122, 222\}$. $\hat{\tau}$, $\hat{\tau}_{[1]}$, and $\hat{\tau}_{[2]}$ are the point estimators, and the numbers in the parentheses are the estimated standard errors. The bold numbers correspond to those peer effects significantly different from zero at level 0.05.

Department	Estimator	(r_1, r_2)	(r_1, r_3)	(r_1, r_4)	(r_2, r_3)	(r_2, r_4)	(r_3, r_4)
Informatics	$\hat{\tau}$	−0.117	0.008	**−0.313**	0.125	**−0.196**	**−0.321**
		(0.086)	(0.109)	(0.071)	(0.101)	(0.059)	(0.089)
	$\hat{\tau}_{[1]}$	−0.117	0.074	**−0.285**	0.191	**−0.168**	**−0.359**
		(0.112)	(0.152)	(0.087)	(0.145)	(0.074)	(0.127)
	$\hat{\tau}_{[2]}$	−0.119	−0.125	**−0.369**	−0.006	**−0.250**	**−0.244**
		(0.126)	(0.120)	(0.123)	(0.092)	(0.096)	(0.088)
Physics	$\hat{\tau}$	0.172	−0.108	−0.095	**−0.280**	−0.267	0.013
		(0.142)	(0.140)	(0.177)	(0.103)	(0.150)	(0.148)
	$\hat{\tau}_{[1]}$	0.329	−0.099	0.017	**−0.427**	−0.311	0.116
		(0.185)	(0.167)	(0.268)	(0.145)	(0.255)	(0.243)
	$\hat{\tau}_{[2]}$	−0.007	−0.119	−0.222	−0.112	−0.215	−0.103
		(0.219)	(0.231)	(0.225)	(0.145)	(0.135)	(0.153)

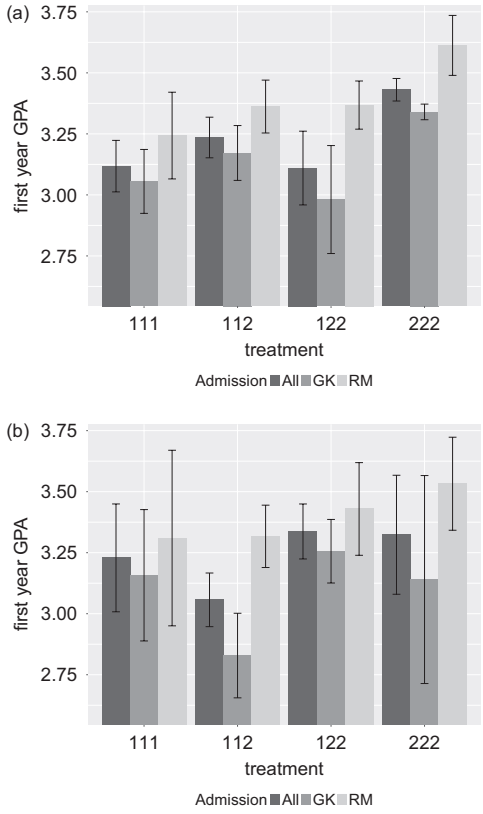


Figure 2. Estimated average and subgroup average potential outcomes with $\{r_1, r_2, r_3, r_4\} = \{111, 112, 122, 222\}$. The black, gray, and light-gray bars denote the estimated average potential outcomes under different treatments for all students, students admitted through Gaokao and recommendation, respectively, and the solid lines denote the 95% confidence intervals. (a) Informatics (b) Physics.

sets, the treatment assignment mechanism is equivalent to complete randomization. Recall that students from Gaokao have attribute 1, and students from recommendation have attribute 2. Under Assumptions 1 and 2, we have four treatments, contained in $\mathcal{R} = \{r_1, r_2, r_3, r_4\} = \{111, 112, 122, 222\}$.

Table 2 shows the estimated average peer effects for these two departments with estimated standard errors based on Corollary 3. Treatment r_4 is significantly better than other treatments for students in the informatics department. Treatment r_3 is significantly better than treatment r_2 for students from Gaokao in the physics department.

Figure 2 shows the estimated average potential outcomes $\hat{Y}(r)$ and $\hat{Y}_{[a]}(r)$, as well as their 95% confidence intervals for all possible a and r. It displays some interesting results. Students from recommendation in both departments have higher average GPAs if they have more roommates from recommendation. However, this monotonic pattern does not apply to students from Gaokao: in the informatics department, the average GPA drops when the number of their peers from recommendation increases from 1 to 2 (i.e., the treatment moves from $r_2 = 112$ to $r_3 = 122$); in the physics department, the average GPA drops when the number of their peers from recommendation increases from 0 to 1 (i.e., the treatment moves from $r_1 = 111$ to $r_2 = 112$) and drops again when the number of their peers from recommendation increases from 2 to 3 (i.e., the treatment moves from $r_3 = 122$ to $r_4 = 222$). We observe some treatment effect heterogeneity in different subgroups, although many results in Table 2 are insignificant due to small sample sizes.

6.3. Optimal Roommate Assignment

We derive the optimal roommate assignment mechanism for the same population as those in the informatics or physics departments, separately. We estimate the optimal complete randomization and obtain the fiducial distribution of l_{opt}. Table 3 shows the estimators and fiducial distributions for the optimal roommate assignments.

As Table 3 suggests, assigning students with the same type together can maximize the academic performance of all students. This may encourage separating students admitted through different channels. However, the optimal treatment assignment has huge uncertainty. The decision may be misleading based on a point estimate of the optimal treatment assignment. It is worth taking a look at the optimal treatment assignments with the top five fiducial probabilities. Most of them do suggest mixing students with different attributes. Moreover, in practice, the average GPA is just a single measure of the students' performance, and other criteria may come into play in practice. For example, if we consider both the average GPA and the diversity of students in each room, it is better to mix different types of students.

6.4. Future Research Directions Based on this Dataset

There are several interesting future research directions based on this dataset. First, we used the largest two departments for randomization-based inference, because large-sample approxi-

Table 3. Fiducial distributions of optimal roommate assignments. Estimated optimal roommate assignments are in bold with $\mathcal{G} = \{g_1, \ldots, g_5\} = \{1111, 1112, 1122, 1222, 2222\}$. The "Prob." columns denote the fiducial probability of the treatment assignment based on 10^4 draws, and the "Outcome" columns denote the unbiased estimator for the expected total GPA under any treatment assignment mechanism.

Informatics							Physics						
Prob.	Outcome	l_1	l_2	l_3	l_4	l_5	Prob.	Outcome	l_1	l_2	l_3	l_4	l_5
0.488	505.60	**26**	**0**	**0**	**0**	**13**	0.555	306.29	**12**	**0**	**0**	**1**	**10**
0.209	504.96	2	32	0	0	5	0.160	301.61	2	0	20	1	0
0.159	504.27	0	34	1	0	4	0.132	302.36	9	0	0	13	1
0.091	504.03	0	34	0	2	3	0.059	300.49	1	1	21	0	0
0.015	502.70	0	33	0	5	1	0.049	301.89	8	0	2	13	0
0.009	496.64	0	26	13	0	0	0.022	305.17	11	1	1	0	10
0.009	488.71	13	0	26	0	0	0.010	300.82	8	1	0	14	0
0.008	498.42	22	0	0	16	1	0.004	288.48	1	15	0	0	7
0.007	497.95	21	1	0	17	0	0.004	302.99	10	3	0	0	10
0.003	501.62	0	32	1	6	0	0.003	287.62	0	13	0	10	0
0.002	501.84	1	31	0	7	0	0.002	298.31	0	3	20	0	0

mations for other small departments are unlikely to be reliable. Second, we analyzed the data from different departments separately. It would be interesting to analyze the dataset of the whole university simultaneously, allowing the smaller departments to borrow information from other larger departments. Third, the dataset also contains other background information. It is our future research to leverage these covariates to improve estimation efficiency.

7. Discussion: Inference Without Assumptions 1 or 2

Assumptions 1 and 2 may be too strong and may not hold in some applications. Below we discuss alternative inferential strategies without them. We summarize the main results below and relegate the technical details to the supplementary material.

7.1. Randomization Test

Without Assumptions 1 or 2, we can still use the randomization tests under the sharp null hypothesis that the treatment Z does not affect any units. This preserves the Type I error in finite samples. However, rejecting the sharp null hypothesis may not be informative for understanding peer effects. It is worth extending previous randomization test strategies (Rosenbaum 2007; Luo et al. 2012; Aronow 2012; Bowers, Fredrickson, and Panagopoulos 2013; Rigdon and Hudgens 2015; Basse, Feller, and Toulis 2019; Athey, Eckles, and Imbens 2018) to our setting.

7.2. Other Estimands of Interest

We can unbiasedly estimate some other estimands without Assumptions 1 or 2. For example, let $Y_i^{\mathcal{D}}(r) = \sum_{z\in\mathcal{Z}} \text{pr}(Z = z \mid R_i = r) Y_i(z)$ be a weighted average of unit i's potential outcomes, where the superscript $\mathcal{D}$ denotes the design. For example, $\mathcal{D} = \text{RP}$ for random partitioning and $\mathcal{D} = \text{CR}$ for complete randomization. Because the weight is nonzero only for assignment z such that $R_i(z_i) = r$, we can view $Y_i^{\mathcal{D}}(r)$ as a summary of the potential outcomes $Y_i(z)$'s when unit i has K peers with attributes r. Moreover, if Assumption 1 holds, $Y_i^{\text{RP}}(r) = Y_i^{\text{CR}}(r)$ reduces to the average of the $Y_i(z_i)$'s for z_i such that $R_i(z_i) = r$. Thus, we can view $\tau_i^{\mathcal{D}}(r, r') = Y_i^{\mathcal{D}}(r) - Y_i^{\mathcal{D}}(r')$ as an individual peer effect comparing treatments r and r'. Define $\bar{Y}_{[a]}^{\mathcal{D}}(r)$ and $\tau_{[a]}^{\mathcal{D}}(r, r')$ as the averages of $Y_i^{\mathcal{D}}(r)$'s and $\tau_i^{\mathcal{D}}(r, r')$'s for units with attribute a, and $\tau^{\mathcal{D}}(r, r')$ as the average of $\tau_i^{\mathcal{D}}(r, r')$'s for all units. These estimands depend on the design as emphasized by the superscript $\mathcal{D}$. In contrast, the estimands $\tau_{[a]}(r, r')$ and $\tau(r, r')$ in previous sections do not depend on the design. Under treatment assignment mechanisms satisfying Assumption 3, we can show that the estimators $\hat{Y}_{[a]}(r)$, $\hat{\tau}_{[a]}(r, r')$, and $\hat{\tau}(r, r')$ in (5) and (6) are still unbiased for $\bar{Y}_{[a]}^{\mathcal{D}}(r)$, $\tau_{[a]}^{\mathcal{D}}(r, r')$, and $\tau^{\mathcal{D}}(r, r')$, respectively. However, we do not have replications for any treatment levels to evaluate their uncertainty.

In sum, Assumptions 1 and 2 can be strong in practice. Without them, it is challenging to conduct repeated sampling inference although it is possible to obtain meaningful point estimates.

7.3. Distributional Assumptions on Potential Outcomes

An alternative approach imposes some distributional assumptions on the potential outcomes. In particular, instead of assuming that the potential outcomes depend only on the peer attribute set as in Assumption 2, we allow for some deviations but need an additional distributional assumption on the error terms.

Assumption 4. The potential outcome can be decomposed as $Y_i(z) = Y_i(R_i(z_i)) + \varepsilon_i(z)$ with the error terms satisfying

(i) $(\varepsilon_1(z), \ldots, \varepsilon_n(z))$ are mutually independent with zero mean, for any peer assignment $z \in \mathcal{Z}$;
(ii) all elements in $\{\varepsilon_i(z) : A_i = a, R_i(z_i) = r, i = 1, \ldots, n, z \in \mathcal{Z}\}$ have the same variance $\sigma_{[a]r}^2$, for any attribute $1 \le a \le H$ and any peer attribute set $r \in \mathcal{R}$.

Assumption 4 is weaker than Assumption 2, and reduces to Assumption 2 when the $\varepsilon_i(z)$'s are all zero, that is, $\sigma_{[a]r}^2 = 0$ for all a and r. Under Assumption 4, $\tau_{[a]}(r, r')$ in (1) is still a meaningful estimand, although it depends only on the main terms instead of the potential outcomes. Moreover, it equals the expectation of $\tau_{[a]}^{\mathcal{D}}(r, r')$ defined in Section 7.2 by averaging over the random error terms. Under complete randomization, we show in the supplementary material that $\hat{\tau}_{[a]}(r, r')$ is still an unbiased estimator for $\tau_{[a]}(r, r')$, and the variance estimator

$\hat{V}_{[a]}(r, r')$ is still conservative in expectation for the sampling variance of $\hat{\tau}_{[a]}(r, r')$.

7.4. Peer Effects for a Target Subpopulation

The last strategy is to focus on a smaller subpopulation for which Assumptions 1 and 2 are more plausible.

7.4.1. Target Subpopulation, Potential Outcomes, and Peer Effects

We formulate the approach of Langenskiöld and Rubin (2008) using the potential outcomes introduced in this article. We consider the following ideal setting. First, we select a "target" subpopulation from units with attribute a. We assume that the units in the target subpopulation have identity numbers $(1, 2, \ldots, \underline{m})$ with $\underline{m} \leq \min(m, n_{[a]})$. Second, we assign all units except the target subpopulation into m groups, with $\underline{m} \leq m$ groups containing K units and the remaining $m - \underline{m}$ groups containing $K+1$ units. Third, we assign the $\underline{m}$ units in the target subpopulation into these $\underline{m}$ groups with K units. Therefore, the peer assignments for units in the target subpopulation are from randomly permuting these $\underline{m}$ groups.

Let $(\zeta_1, \ldots, \zeta_{\underline{m}})$ denote the units initially assigned to the $\underline{m}$ groups, where ζ_k is the set consisting of the identity numbers of the K units in group k $(1 \leq k \leq \underline{m})$. Therefore, the peer assignments for units in the target subpopulation, $(z_1, \ldots, z_{\underline{m}})$, is a permutation of $(\zeta_1, \ldots, \zeta_{\underline{m}})$, and the peer assignments for units in the remaining $m - \underline{m}$ groups are fixed. Therefore, unit i's potential outcome simplifies to $Y_i(z_1, \ldots, z_{\underline{m}}, z_{\underline{m}+1}, \ldots, z_n) = Y_i(z_1, \ldots, z_{\underline{m}})$ for $1 \leq i \leq \underline{m}$.

Following Langenskiöld and Rubin (2008), we introduce the following two assumptions.

Assumption 5. If $z_i = z_i'$, then $Y_i(z_1, \ldots, z_{\underline{m}}) = Y_i(z_1', \ldots, z_{\underline{m}}')$, for any two peer assignments $(z_1, \ldots, z_{\underline{m}})$ and $(z_1', \ldots, z_{\underline{m}}')$ and for any unit $1 \leq i \leq \underline{m}$ in the target subpopulation.

Assumption 5 requires that each unit's potential outcomes depend only on its own peers. Under Assumption 5, unit i's potential outcome simplifies to $Y_i(z_1, \ldots, z_{\underline{m}}) = Y_i(z_i)$ for $1 \leq i \leq \underline{m}$.

Assumption 6. If $R_i(z_i) = R_i(z_i')$, then $Y_i(z_i) = Y_i(z_i')$, for any two peer assignments $(z_1, \ldots, z_{\underline{m}})$ and $(z_1', \ldots, z_{\underline{m}}')$ and for any unit $1 \leq i \leq \underline{m}$ in the target subpopulation.

Assumption 6 requires that each unit's potential outcomes depend only on the attributes of its peers. Under Assumption 6, unit i's potential outcome simplifies to $Y_i(z_i) = Y_i(R_i(z_i))$ for $1 \leq i \leq \underline{m}$.

Langenskiöld and Rubin (2008) chose the attribute to be the smoking behavior and the target subpopulation to be nonsmoking freshman. They further dichotomized each of $(\zeta_1, \ldots, \zeta_{\underline{m}})$ into two categories: smoking or nonsmoking suites.

Under Assumptions 5 and 6, unit i's potential outcome simplifies to $Y_i(r)$ for some $r \in \mathcal{R}$, for $1 \leq i \leq \underline{m}$ in the target subpopulation. Comparing two treatments $r, r' \in \mathcal{R}$, the individual peer effect for unit i is $\tau_i(r, r') = Y_i(r) - Y_i(r')$, and the average peer effect for units in the target subpopulation is $\tau_{\text{tg}}(r, r') = \underline{m}^{-1} \sum_{i=1}^{\underline{m}} \tau_i(r, r')$. We want to infer τ_{tg}.

7.4.2. Construction of Target Subpopulation and Statistical Inference

We first construct the target subpopulation and then infer the peer effects for it. In the following, we assume complete randomization. For the observed peer assignment Z, let $\underline{m}$ be the number of groups with units of attribute a. For each of the $\underline{m}$ groups, we randomly pick one unit with attribute a to constitute the target subpopulation, and denote the remaining units in these groups as $(\zeta_1, \ldots, \zeta_{\underline{m}})$. To construct the ideal setting, we conduct inference conditional on the group assignments for all units excluding the target subpopulation. The remaining randomness comes solely from the peer assignments of the target subpopulation, which is a random permutation of $(\zeta_1, \ldots, \zeta_{\underline{m}})$. Moreover, under Assumptions 5 and 6, the treatment assignment is a completely randomized experiment with multiple treatments taking values in $\mathcal{R} = \{r_1, \ldots, r_{|\mathcal{R}|}\}$. Therefore, for the average peer effect $\tau_{\text{tg}}(r, r')$, an unbiased estimator is the standard difference-in-means for units receiving treatments r and r'. We relegate the sampling variance and variance estimator to the supplementary material.

7.4.3. Comparison and Connection to Our Approach

Compared to Assumptions 1 and 2, Assumptions 5 and 6 are weaker because they make assumptions for a subset of units and a subset of peer assignments, that is, the unique values in $(\zeta_1, \ldots, \zeta_{\underline{m}})$. However, under this ideal setting with Assumptions 5 and 6, we can only infer peer effects for the target subpopulation instead of all units.

The construction in Section 7.4.2 generates a random target subpopulation. Interestingly, averaging over all possible constructions, the point estimator for $\tau_{\text{tg}}(r, r')$ is the same as $\hat{\tau}_{[a]}(r, r')$ in (13). We prove this result in the supplementary material.

Supplementary Material

Appendix A1 gives supporting materials for Section 7. Appendix A2 gives more technical details for general treatment assignment mechanisms. Appendix A3 gives more technical details for complete randomization. Appendix A4 gives more technical details for random partitioning.

Acknowledgments

The authors thank Don Rubin, Luke Miratrix, Zach Branson and Kristen Hunter at Harvard University, the associate editor, and two reviewers for insightful comments. Dr. Avi Feller at UC Berkeley kindly edited their final version.

Funding

The authors gratefully acknowledge financial support from the National Science Foundation (Peng Ding: DMS grant no. 1713152; Jun Liu: DMS no. 1712714).

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017), "When Should You Adjust Standard Errors for Clustering?" arXiv: 1710.02926 [1658]

An, W. (2011), "Models and Methods to Identify Peer Effects," in *The Sage Handbook of Social Network Analysis*, eds. J. Scott and P. J. Carrington, London: Sage, pp. 515–532. [1651]

Angrist, J. D. (2014), "The Perils of Peer Effects," *Labour Economics*, 30, 98–108. [1658]

Aronow, P. M. (2012), "A General method for Detecting Interference between units in Randomized Experiments," *Sociological Methods and Research*, 41, 3–16. [1652,1655,1662]

Aronow, P. M., and Samii, C. (2017), "Estimating Average Causal Effects under Interference between Units," *Annals of Applied Statistics*, 11, 1912–1947. [1652]

Arpino, B., and Mattei, A. (2016), "Assessing the Causal Effects of Financial Aids to Firms in Tuscany Allowing for Interference," *Annals of Applied Statistics*, 10, 1170–1194. [1651]

Athey, S., Eckles, D., and Imbens, G. W. (2018), "Exact P-values for Network Interference," *Journal of the American Statistical Association*, 113, 230–240. [1652,1655,1662]

Basse, G., and Feller, A. (2018), "Analyzing Two-Stage Experiments in the Presence of Interference," *Journal of the American Statistical Association*, 113, 41–55. [1652]

Basse, G., Feller, A., and Toulis, P. (2019), "Randomization tests of causal effects under interference," *Biometrika*, doi: 10.1093/biomet/asy072 [1652,1662]

Bhattacharya, D. (2009), "Inferring Optimal Peer Assignment from Experimental Data," *Journal of the American Statistical Association*, 104, 486–500. [1651,1660]

Bowers, J., Fredrickson, M. M., and Panagopoulos, C. (2013), "Reasoning about Interference between Units: A General Framework," *Political Analysis*, 21, 97–124. [1651,1652,1662]

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013), "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81, 855–882. [1658]

Choi, D. S. (2017), "Estimation of Monotone Treatment Effects in Network Experiments," *Journal of the American Statistical Association*, 112, 1147–1155. [1652]

Cox, D. R. (1958), *Planning of Experiments*, Oxford, England: Wiley. [1651]

Dasgupta, T., Pillai, N. S., and Rubin, D. B. (2015), "Causal Inference from 2^K Factorial Designs by using Potential Outcomes," *Journal of the Royal Statistical Society*, Series B, 77, 727–753. [1660]

Fisher, R. A. (1935), *The Design of Experiments* (1st ed.), Edinburgh, London: Oliver and Boyd. [1660]

Forastiere, L., Airoldi, E. M., and Mealli, F. (2016), "Identification and Estimation of Treatment and Interference effects in Observational Studies on Networks," arXiv:1609.06245. [1651]

Goldsmith-Pinkham, P., and Imbens, G. W. (2013), "Social Networks and the Identification of Peer effects," *Journal of Business and Economic Statistics*, 31, 253–264. [1651]

Graham, B. S., Imbens, G. W., and Ridder, G. (2010), "Measuring the Effects of Segregation in the Presence of Social Spillovers: A Nonparametric Approach," National Bureau of Economic Research Working Paper 16499, National Bureau of Economic Research. [1651]

Halloran, M. E., and Struchiner, C. J. (1991), "Study Designs for Dependent Happenings," *Epidemiology*, 2, 331–338. [1651]

——— (1995), "Causal Inference in Infectious Diseases," *Epidemiology*, 6, 142–151. [1651]

Hong, G., and Raudenbush, S. W. (2006), "Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data," *Journal of the American Statistical Association*, 101, 901–910. [1651]

Hudgens, M. G., and Halloran, M. E. (2008), "Toward Causal Inference With Interference," *Journal of the American Statistical Association*, 103, 832–842. [1652,1653,1655]

Ichino, N., and Schündeln, M. (2012), "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana," *The Journal of Politics*, 74, 292–307. [1651]

Kwan, Y. K. (1999), "Asymptotic Bayesian Analysis based on a Limited Information Estimator," *Journal of Econometrics*, 88, 99–121. [1660]

Langenskiöld, S., and Rubin, D. B. (2008), "Outcome-free Design of Observational Studies: Peer Influence on Smoking," *Annales d'Économie et de Statistique*, 91–92, 107–125. [1653,1663]

Li, X., and Ding, P. (2017), "General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference," *Journal of the American Statistical Association*, 112, 1759–1769. [1657,1658]

Lin, W. (2013), "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *The Annals of Applied Statistics*, 7, 295–318. [1658]

Liu, L., and Hudgens, M. G. (2014), "Large Sample Randomization Inference of Causal effects in the Presence of Interference," *Journal of the American Statistical Association*, 109, 288–301. [1652,1655]

Liu, L., Hudgens, M. G., and Becker-Dreps, S. (2016), "On Inverse Probability-Weighted Estimators in the Presence of Interference," *Biometrika*, 103, 829–842. [1652]

Luo, X., Small, D. S., Li, C. S. R., and Rosenbaum, P. R. (2012), "Inference with Interference between Units in an fMRI Experiment of Motor Inhibition," *Journal of the American Statistical Association*, 107, 530–541. [1652,1662]

Manski, C. F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531–542. [1651,1658]

——— (2013), "Identification of Treatment Response with Social Interactions," *The Econometrics Journal*, 16, S1–S23. [1653]

Miguel, E., and Kremer, M. (2004), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72, 159–217. [1651]

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles" (with discussion), Section 9 (translated), reprinted ed., *Statistical Science*, 5, 465–472. [1651,1656,1657]

Nickerson, D. W. (2008), "Is Voting Contagious? Evidence from two Field Experiments," *American Political Science Review*, 102, 49–57. [1651]

Ogburn, E. L., and VanderWeele, T. J. (2014), "Causal Diagrams for Interference," *Statistical Science*, 29, 559–578. [1651]

Perez-Heydrich, C., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M., and Emch, M. E. (2014), "Assessing Effects of Cholera Vaccination in the Presence of Interference," *Biometrics*, 70, 731–741. [1651,1652]

Rigdon, J., and Hudgens, M. G. (2015), "Exact Confidence Intervals in the Presence of Interference," *Statistics and Probability Letters*, 105, 130–135. [1652,1662]

Rosenbaum, P. R. (2007), "Interference between units in Randomized Experiments," *Journal of the American Statistical Association*, 102, 191–200. [1652,1662]

Rubin, D. B. (1980), "Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization test" by D. Basu," *Journal of American Statistical Association*, 75, 591–593. [1651]

——— (2005), "Causal Inference using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association*, 100, 322–331. [1652]

Sacerdote, B. (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 116, 681–704. [1651,1652,1653,1658]

Sävje, F., Aronow, P. M., and Hudgens, M. G. (2017), "Average Treatment Effects in the Presence of Unknown Interference," arXiv:1711.06399. [1652]

Sims, C. A. (2006), "On An Example of Larry Wasserman," Technical Report, Princeton University, available at *http://sims.princeton.edu/yftp/WassermanExmpl/WassermanComment.pdf* [1660]

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, 1398–1407. [1652,1653]

Tchetgen, E. J. T., and VanderWeele, T. J. (2012), "On Causal Inference in the Presence of Interference," *Statistical Methods in Medical Research*, 21, 55–75. [1652]

Toulis, P., and Kao, E. (2013), "Estimation of Causal Peer Influence Effects," in *Proceedings of the 30th International Conference on Machine Learning*, pp. 1489–1497. [1652]

VanderWeele, T. J., and An, W. (2013), "Social Networks and Causal Inference," in *Handbook of Causal Analysis for Social Research*, ed., S. L. Morgan, Netherlands: Springer, pp. 353–374. [1651]

Vanderweele, T. J., Hong, G., Jones, S. M., and Brown, J. L. (2013), "Mediation and Spillover Effects in Group-Randomized Trials: A Case Study of the 4Rs Educational Intervention," *Journal of the American Statistical Association*, 108, 469–482. [1651]

Supplementary Material for "Randomization Inference for Peer Effects"

by Xinran Li, Peng Ding, Qian Lin, Dawei Yang, and Jun Liu

Appendix A1 gives supporting materials for Section 7.
Appendix A2 gives more technical details for general treatment assignment mechanisms.
Appendix A3 gives more technical details for complete randomization.
Appendix A4 gives more technical details for random partitioning.

A1. Analyzing peer effects without Assumptions 1 or 2

A1.1. Randomization tests

We use randomization tests for the significance of peer effects for the following reasons. First, they provide additional evidence for the significance of peer effects. Second, randomization tests are exact and valid for finite samples. Third, randomization tests do not require Assumptions 1–3, as long as the assignment mechanism is known. Fourth, as Fisher (1935) suggested, we can use randomization tests to check the Normal approximations.

We can use randomization tests for the sharp null hypothesis for all units:

$$H_0 : Y_i(z) = Y_i(z'), \quad \text{for all } z, z' \text{ and for all unit } i,$$

or the null hypothesis for units with attribute a:

$$H_{0,[a]} : Y_i(z) = Y_i(z'), \quad \text{for all } z, z' \text{ and for all unit } i \text{ such that } A_i = a.$$

$H_{0,[a]}$ is not sharp. But we can still conduct randomization test for $H_{0,[a]}$. We choose test statistics depending only on the outcomes of units with attribute a, and their randomization distributions are known under $H_{0,[a]}$. We can then obtain exact p-values under $H_{0,[a]}$.

We first discuss the choices of test statistics for the subgroup null $H_{0,[a]}$, and then the test statistics for H_0. A choice of test statistic for the subgroup null $H_{0,[a]}$ is

$$T_{[a]} = \max_{r,r'} \hat{\tau}_{[a]}(r, r') = \max_r \hat{Y}_{[a]}(r) - \min_r \hat{Y}_{[a]}(r). \tag{A1}$$

Another choice of test statistic, $F_{[a]}$, is the F statistic from the analysis of variance of the linear regression of Y_i on R_i among units with attribute a. For the sharp null H_0, we can use $T = \max_{r,r'} \hat{\tau}(r, r')$, the F statistic F from the linear model (16), $\max_a T_{[a]}$, and $\max_a F_{[a]}$.

For the motivating application, Table A1 shows the p-values from randomization tests for H_0 and $H_{0,[a]}$ with different test statistics. At significance level 0.05, the peer effects are not significant

Table A1: p values of randomization tests for the subgroup null $H_{0,[a]}$ and the sharp null H_0, with test statistics for H_0 shown in the parentheses.

Department	Null hypothesis	Test statistic for $H_{0,[a]}$ (H_0)			
		$T_{[a]}$ ($\max_a T_{[a]}$)	$F_{[a]}$ ($\max_a F_{[a]}$)	(T)	(F)
Informatics	H_0	0.2678	0.3640	0.1092	0.1468
	$H_{0,[1]}$	0.2390	0.3384		
	$H_{0,[2]}$	0.0615	0.2050		
physics	H_0	0.4553	0.0719	0.3431	**0.0366**
	$H_{0,[1]}$	0.3545	**0.0360**		
	$H_{0,[2]}$	0.6994	0.7232		

for students in the informatics department; the subgroup average peer effects are significant for students from Gaokao in the physics department if we use the F statistic as the test statistic, and the subgroup average peer effects are not significant for students from recommendation. We ignore the multiple testing issue. Compared to Table 2, randomization tests reject only $H_{0,[1]}$ for students from Gaokao in the physics department, which may be due to the lack of power of randomization tests (Ding 2017).

Moreover, we check the randomization distributions of the $\hat{\tau}_{[a]}(r, r')$'s under the sharp null hypothesis. Under the sharp null hypothesis that the potential outcomes are not affected by the treatment assignment, all potential outcomes are known and identical to the observed outcomes. Therefore, the distributions of the subgroup peer effect estimators are known under complete randomization. For students in the informatics and physics departments graduating in 2013, Figures A1(a) and A1(b) show, respectively, the histograms of the subgroup peer effect estimators under the sharp null hypothesis based on 10^5 treatment assignments from complete randomization. From Figures A1(a) and A1(b), the Normal approximations work fairly well. For our application, we do not know all potential outcomes and thus can not directly check the Normal approximations over repeated sampling of the treatment assignments. However, we view Figures A1(a) and A1(b) as intuitive justifications for the Normal approximation of $\hat{\tau}_{[a]}(r, r')$ in the Neymanian inference.

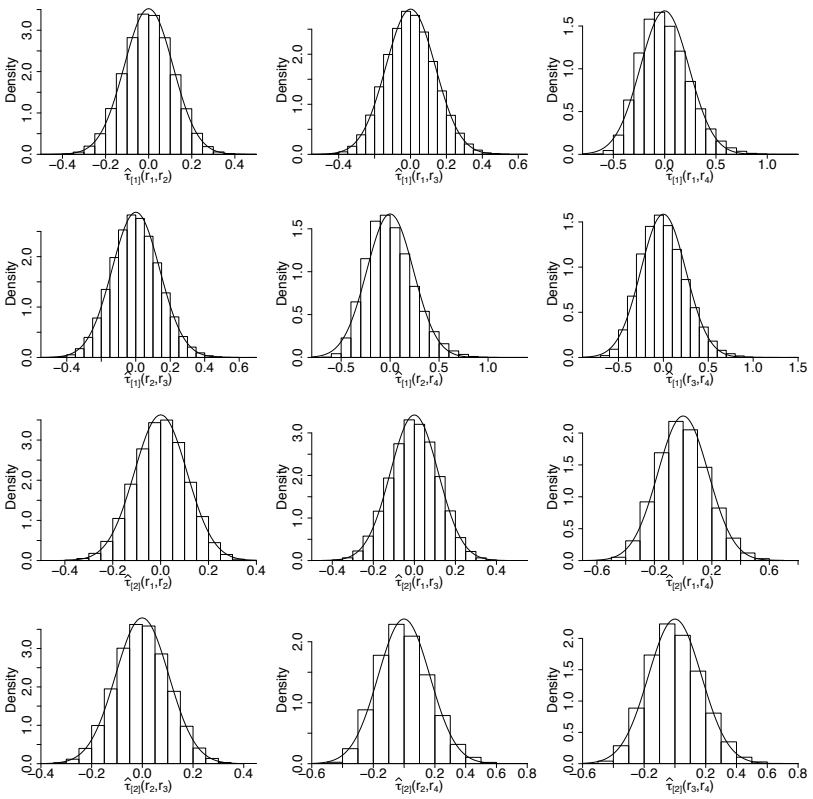
A1.2. Estimands, unbiased estimators, and optimal assignment mechanism

Even if Assumptions 1 or 2 fails, the estimators in (5) and (6) are still meaningful in the sense of unbiasedly estimating some average potential outcomes. Recall the definitions in Section 7:

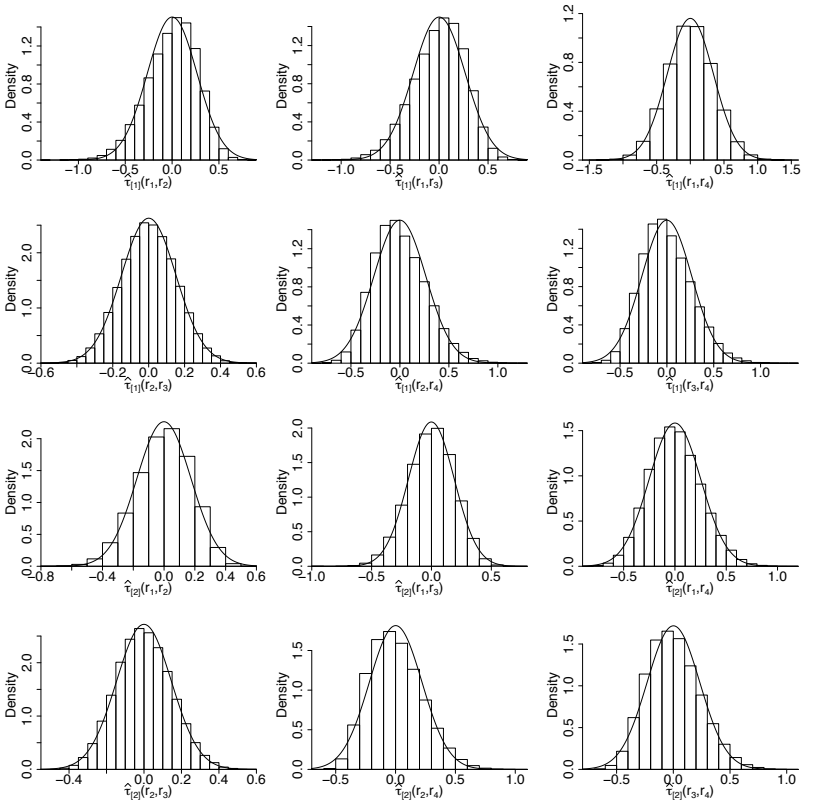
$$Y_i^{\mathcal{D}}(r) = \sum_{z \in \mathcal{Z}} \text{pr}(Z = z \mid R_i = r) Y_i(z), \quad \tau_i^{\mathcal{D}}(r, r') = Y_i^{\mathcal{D}}(r) - Y_i^{\mathcal{D}}(r'),$$

$$\bar{Y}_{[a]}^{\mathcal{D}}(r) = n_{[a]}^{-1} \sum_{i: A_i = a} Y_i^{\mathcal{D}}(r), \quad \tau_{[a]}^{\mathcal{D}}(r, r') = n_{[a]}^{-1} \sum_{i: A_i = a} \tau_i^{\mathcal{D}}(r, r'), \quad \tau^{\mathcal{D}}(r, r') = n^{-1} \sum_{i=1}^{n} \tau_i^{\mathcal{D}}(r, r').$$

Proposition A1. Under Assumption 3, $\hat{Y}_{[a]}(r)$ has mean $\bar{Y}_{[a]}^{\mathcal{D}}(r)$.



(a) The informatics department



(b) The physics department

Figure A1: Histograms of subgroup peer effect estimators under the sharp null hypothesis, based on 10^5 draws from complete randomization. The lines are densities of Normal approximations.

The conclusion follows from

$$
\begin{aligned}
E\left\{\hat{Y}_{[a]}(r)\right\} &= \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} E\left\{I(R_i=r)Y_i\right\} \\
&= \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} E\left\{I(R_i=r)\sum_z I(Z=z)Y_i(z)\right\} \\
&= \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a}\sum_z E\left\{I(R_i=r)I(Z=z)\right\}Y_i(z) \\
&= \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a}\sum_z \mathrm{pr}(R_i=r,Z=z)Y_i(z) \\
&= \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a}\sum_z \mathrm{pr}(R_i=r)\mathrm{pr}(Z=z\mid R_i=r)Y_i(z) \\
&= n_{[a]}^{-1} \sum_{i:A_i=a}\sum_z \mathrm{pr}(Z=z\mid R_i=r)Y_i(z) = n_{[a]}^{-1}\sum_{i:A_i=a} Y_i^{\mathcal{D}}(r) = \bar{Y}_{[a]}^{\mathcal{D}}(r).
\end{aligned}
$$

By the linearity of the expectation, $\hat{\tau}_{[a]}(r,r')$ and $\hat{\tau}(r,r')$ are unbiased for $\tau_{[a]}^{\mathcal{D}}(r,r')$ and $\tau^{\mathcal{D}}(r,r')$, respectively. However, without Assumptions 1 and 2, for a given treatment we do not have replications of units, making it difficult to evaluate the uncertainty of these estimators. Similarly, for the first type of interference, Hudgens and Halloran (2008) discussed the expectations of the point estimators under general settings, but invoked "stratified interference" (analogous to Assumption 2) to evaluate the uncertainty.

Moreover, the estimands $\tau_{[a]}^{\mathcal{D}}(r,r')$ and $\tau^{\mathcal{D}}(r,r')$ are meaningful in many situations. In the expression of $Y_i^{\mathcal{D}}(r)$, the weight $\mathrm{pr}(Z=z\mid R_i=r)$ is nonzero only if $R_i(z_i)=r$, i.e., the attributes of unit i's peers constitute r. Therefore, $Y_i^{\mathcal{D}}(r)$ summarizes unit i's potential outcomes when he/she has K peers with attributes r. Consequently, $\tau_i^{\mathcal{D}}(r,r')$ measures the difference when unit i has K peers with attributes r rather than r'. Thus we can view $\tau_i^{\mathcal{D}}(r,r')$ as the individual peer effect comparing treatments r and r', and $\tau_{[a]}^{\mathcal{D}}(r,r')$ and $\tau^{\mathcal{D}}(r,r')$ as the corresponding average peer effects. Below we further simplify $Y_i^{\mathcal{D}}(r)$ under some special cases, making its meaning more intuitive. When Assumption 1 holds, $Y_i^{\mathcal{D}}(r)$ reduces to $\sum_{z_i}\mathrm{pr}(Z_i=z_i\mid R_i=r)Y_i(z_i)$; if further the treatment assignment mechanism is random partitioning or complete randomization, then the weight $\mathrm{pr}(Z_i=z_i\mid R_i=r)$ is a nonzero constant for z_i such that $R_i(z_i)=r$, and $Y_i^{\mathrm{RP}}(r)=Y_i^{\mathrm{CR}}(r)$ further reduces to the average of $Y_i(z_i)$'s for z_i such that $R_i(z_i)=r$. When both Assumptions 1 and 2 hold, $Y_i^{\mathcal{D}}(r)$ is the same as $Y_i(r)$ in the main paper, which does not depend on the assignment mechanism $\mathcal{D}$. In sum, $Y_i^{\mathcal{D}}(r)$ is an extension of $Y_i(r)$.

We now consider the optimal complete randomization mechanism for the assignment of a new population of size n' without Assumptions 1 or 2. Define similarly $Y_i^{\mathcal{D}'}(r)$ and $\bar{Y}_{[a]}^{\mathcal{D}'}(r)$ for the new population. By the same logic as (18), the expected total outcome under complete randomization with $L'(z)=l'$ for the new population is

$$
E\left(\sum_{i=1}^{n'} Y_i'\right) = \sum_{a=1}^{H} E\left(\sum_{i:A_i=a} Y_i'\right) = \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n_{[a]r}'\bar{Y}_{[a]}^{\mathrm{CR}'}(r). \tag{A2}
$$

To estimate the maximizer l'_{opt} of (A2), we need to assume the new population is similar to the one in our data. Let $\mathcal{D}_0$ denote the assignment mechanism for our observed data. The following assumption is similar to the one in the main paper.

Assumption A7. $\bar{Y}^{\text{CR}'}_{[a]}(r) = c\bar{Y}^{\mathcal{D}_0}_{[a]}(r) + \xi_{[a]}$ for some constants $c > 0$ and $\xi_{[a]}$, for all $1 \le a \le H$ and $r \in \mathcal{R}$.

Under Assumption A7, (A2) reduces to

$$E\left(\sum_{i=1}^{n'} Y'_i\right) = \sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}^{\text{CR}'}_{[a]}(r) = c\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}^{\mathcal{D}_0}_{[a]}(r) + \sum_{a=1}^{H} n'_{[a]}\xi_{[a]},$$

implying that the maximizer l'_{opt} of (A2) is the same as that of $\sum_{a=1}^{H}\sum_{r\in\mathcal{R}} n'_{[a]r}\bar{Y}^{\mathcal{D}_0}_{[a]}(r)$. We can unbiasedly estimate $\bar{Y}^{\mathcal{D}_0}_{[a]}(r)$ by $\hat{Y}_{[a]}(r)$, and then estimate l'_{opt} by simply plugging in the estimators $\hat{Y}_{[a]}(r)$'s. Again, it is difficult to evaluate the uncertainty of the estimator for l'_{opt} for the same reason as that for the average peer effect estimators.

Below we give some comments on Assumption A7, which is key for inferring the optimal complete randomization. Assumption A7 is a strong requirement of the similarity between the new population and the one in our data, due to the dependence of $\bar{Y}^{\text{CR}'}_{[a]}(r)$ and $\bar{Y}^{\mathcal{D}_0}_{[a]}(r)$ on the designs. Even if we assume that the new population is the same as the one in our data, Assumption A7, or, equivalently, $\bar{Y}^{\text{CR}}_{[a]}(r) = c\bar{Y}^{\mathcal{D}_0}_{[a]}(r) + \xi_{[a]}$, may fail because the values of $\bar{Y}^{\text{CR}}_{[a]}(r)$ and $\bar{Y}^{\mathcal{D}_0}_{[a]}(r)$ depend on the assignment mechanisms, CR and $\mathcal{D}_0$, respectively. If the new population is different from the one in our data, then Assumption A7 is even less plausible.

We summarize several concerns for inferring the optimal complete randomization in the absence of Assumptions 1 or 2. First, it is unnatural to infer the optimal peer assignment for a new population, because the treatment is the set of the identity numbers of units varying across populations. Second, the similarity assumption becomes stronger due to the dependence of the estimands on the design. Third, it is difficult to evaluate the uncertainty of the point estimator.

A1.3. Distributional assumptions on potential outcomes

Under Assumption 4, we decompose $\hat{Y}_{[a]}(r)$ into two parts:

$$\begin{aligned}\hat{Y}_{[a]}(r) &= n^{-1}_{[a]r}\sum_{i:A_i=a,R_i=r} Y_i = n^{-1}_{[a]r}\sum_{i:A_i=a,R_i=r}\{Y_i(r) + \varepsilon_i(Z)\}\\ &= n^{-1}_{[a]r}\sum_{i:A_i=a,R_i=r} Y_i(r) + n^{-1}_{[a]r}\sum_{i:A_i=a,R_i=r}\varepsilon_i(Z) \equiv \check{Y}_{[a]}(r) + \check{\varepsilon}_{[a]}(r),\end{aligned}$$

where $\check{Y}_{[a]}(r) \equiv n^{-1}_{[a]r}\sum_{i:A_i=a,R_i=r} Y_i(r)$ and $\check{\varepsilon}_{[a]}(r) \equiv n^{-1}_{[a]r}\sum_{i:A_i=a,R_i=r}\varepsilon_i(Z)$ are the main part and deviance part, respectively. Let $\check{\tau}_{[a]}(r,r') = \check{Y}_{[a]}(r) - \check{Y}_{[a]}(r')$ and $\check{\delta}_{[a]}(r,r') = \check{\varepsilon}_{[a]}(r) - \check{\varepsilon}_{[a]}(r')$. Correspondingly, we can decompose the subgroup peer effect estimator $\hat{\tau}_{[a]}(r,r')$ into two parts: $\hat{\tau}_{[a]}(r,r') = \check{\tau}_{[a]}(r,r') + \check{\delta}_{[a]}(r,r')$.

First, we discuss the sampling mean and variance of $\hat{\tau}_{[a]}(r, r')$. Under Assumption 4, we have, for any $1 \leq a \leq H$ and $r \neq r' \in \mathcal{R}$,

$$
\begin{aligned}
E\left\{\check{\varepsilon}_{[a]}(r) \mid Z\right\} &= E\left\{n_{[a]r}^{-1} \sum_{i:A_i=a,R_i=r} \varepsilon_i(Z) \mid Z\right\} = n_{[a]r}^{-1} \sum_{i:A_i=a,R_i=r} E\left\{\varepsilon_i(Z) \mid Z\right\} = 0, \\
\operatorname{Var}\left\{\check{\varepsilon}_{[a]}(r) \mid Z\right\} &= \operatorname{Var}\left\{n_{[a]r}^{-1} \sum_{i:A_i=a,R_i=r} \varepsilon_i(Z) \mid Z\right\} = n_{[a]r}^{-2} \sum_{i:A_i=a,R_i=r} \operatorname{Var}\left\{\varepsilon_i(Z) \mid Z\right\} = n_{[a]r}^{-1}\sigma_{[a]r}^2, \\
\operatorname{Cov}\left\{\check{\varepsilon}_{[a]}(r), \check{\varepsilon}_{[a]}(r') \mid Z\right\} &= (n_{[a]r}n_{[a]r'})^{-1} \sum_{i:A_i=a,R_i=r} \sum_{j:A_j=a,R_j=r'} E\left\{\varepsilon_i(Z)\varepsilon_j(Z) \mid Z\right\} = 0, \qquad \text{(A3)}
\end{aligned}
$$

which immediately imply that $E\{\check{\delta}_{[a]}(r, r') \mid Z\} = 0$ and $\operatorname{Var}\{\check{\delta}_{[a]}(r, r') \mid Z\} = n_{[a]r}^{-1}\sigma_{[a]r}^2 + n_{[a]r'}^{-1}\sigma_{[a]r'}^2$. Thus, marginally, $\check{\delta}_{[a]}(r, r')$ has mean 0 and variance $n_{[a]r}^{-1}\sigma_{[a]r}^2 + n_{[a]r'}^{-1}\sigma_{[a]r'}^2$. Because $\check{\tau}_{[a]}(r, r')$ is constant given Z, we have

$$
\operatorname{Cov}\left\{\check{\tau}_{[a]}(r, r'), \check{\delta}_{[a]}(r, r')\right\} = E\left[E\left\{\check{\tau}_{[a]}(r, r')\check{\delta}_{[a]}(r, r') \mid Z\right\}\right] = E\left[\check{\tau}_{[a]}(r, r')E\left\{\check{\delta}_{[a]}(r, r') \mid Z\right\}\right] = 0,
$$

which implies that $\operatorname{Var}\{\hat{\tau}_{[a]}(r, r')\} = \operatorname{Var}\{\check{\tau}_{[a]}(r, r')\} + \operatorname{Var}\{\check{\delta}_{[a]}(r, r')\}$. Because the sampling variance of $\check{\tau}_{[a]}(r, r')$ is the same as that in Corollary 2 with Assumption 2, we can derive that

$$
\begin{aligned}
\operatorname{Var}\left\{\hat{\tau}_{[a]}(r, r')\right\} &= \operatorname{Var}\left\{\check{\tau}_{[a]}(r, r')\right\} + \operatorname{Var}\left\{\check{\delta}_{[a]}(r, r')\right\} = \frac{S_{[a]}^2(r)}{n_{[a]r}} + \frac{S_{[a]}^2(r')}{n_{[a]r'}} - \frac{S_{[a]}^2(r\text{-}r')}{n_{[a]}} + \frac{\sigma_{[a]}^2(r)}{n_{[a]r}} + \frac{\sigma_{[a]}^2(r')}{n_{[a]r'}} \\
&= \frac{S_{[a]}^2(r) + \sigma_{[a]r}^2}{n_{[a]r}} + \frac{S_{[a]}^2(r') + \sigma_{[a]r'}^2}{n_{[a]r'}} - \frac{S_{[a]}^2(r\text{-}r')}{n_{[a]}}.
\end{aligned}
$$

Second, we discuss the variance estimator for $\hat{\tau}_{[a]}(r, r')$. We decompose the sample variance of observed outcomes for units with attribute a receiving treatment r as

$$
\begin{aligned}
s_{[a]}^2(r) &= (n_{[a]r} - 1)^{-1} \sum_{i:A_i=a,R_i=r} \left\{Y_i - \hat{Y}_{[a]}(r)\right\}^2 = (n_{[a]r} - 1)^{-1} \sum_{i:A_i=a,R_i=r} \left\{Y_i(r) + \varepsilon_i(Z) - \check{Y}_{[a]}(r) - \check{\varepsilon}_{[a]}(r)\right\}^2 \\
&= (n_{[a]r} - 1)^{-1} \sum_{i:A_i=a,R_i=r} \left\{Y_i(r) - \check{Y}_{[a]}(r)\right\}^2 + (n_{[a]r} - 1)^{-1} \sum_{i:A_i=a,R_i=r} \left\{\varepsilon_i(Z) - \check{\varepsilon}_{[a]}(r)\right\}^2 \\
&\quad + (n_{[a]r} - 1)^{-1} \sum_{i:A_i=a,R_i=r} \left\{Y_i(r) - \check{Y}_{[a]}(r)\right\}\left\{\varepsilon_i(Z) - \check{\varepsilon}_{[a]}(r)\right\}. \qquad \text{(A4)}
\end{aligned}
$$

Below we discuss the expectation of the three terms in (A4) separately. The expectation of the first term is the same as that in Theorem 2 with Assumption 2. The expectation of the second term is $\sigma_{[a]r'}^2$, because, conditional on Z, it is the sample variance of independent zero-mean random variables with variance $\sigma_{[a]r}^2$. The expectation of the third term is zero, because, conditioning on Z, $Y_i(r) - \check{Y}_{[a]}(r)$ is a constant and $\varepsilon_i(Z) - \check{\varepsilon}_{[a]}(r)$ has mean zero. Above all, $E\{s_{[a]}^2(r)\} = S_{[a]}^2(r) + \sigma_{[a]r}^2$.

The variance estimator is conservative because

$$E\left\{\hat{V}_{[a]}(r,r')\right\} = \frac{E\{s^2_{[a]}(r)\}}{n_{[a]r}} + \frac{E\{s^2_{[a]}(r')\}}{n_{[a]r'}} = \frac{S^2_{[a]}(r)+\sigma^2_{[a]r}}{n_{[a]r}} + \frac{S^2_{[a]}(r')+\sigma^2_{[a]r'}}{n_{[a]r'}} \geq \text{Var}\left\{\hat{\tau}_{[a]}(r,r')\right\}.$$

A1.4. Peer effects for a target subpopulation

First, we study the sampling variance and its estimator for the difference-in-means estimator $\hat{\tau}_{\text{tg}}(r,r')$. For any $r,r' \in \mathcal{R}$ and units in the target subpopulation, let $\bar{Y}_{\text{tg}}(r) = \underline{m}^{-1}\sum_{i=1}^{m} Y_i(r)$ and $\mathfrak{S}^2(r) = (\underline{m}-1)^{-1}\sum_{i=1}^{m}\{Y_i(r)-\bar{Y}_{\text{tg}}(r)\}^2$ be the finite population average and variance of individual potential outcome $Y_i(r)$'s, and $\mathfrak{S}^2(r\text{-}r')$ be the finite population variance of individual peer effect $\tau_i(r,r')$'s. The number of units in the target subpopulation receiving treatment $r \in \mathcal{R}$ is $\underline{m}_r = \sum_{k=1}^{m} I(\{A_j : j \in \zeta_k\} = r)$. Following Neyman (1923), the sampling variance of $\hat{\tau}_{\text{tg}}(r,r')$ is

$$\text{Var}\{\hat{\tau}_{\text{tg}}(r,r')\} = \frac{\mathfrak{S}^2(r)}{\underline{m}_r} + \frac{\mathfrak{S}^2(r')}{\underline{m}_{r'}} - \frac{\mathfrak{S}^2(r\text{-}r')}{\underline{m}}.$$

For any $r \in \mathcal{R}$, let $\mathfrak{s}^2(r)$ be the sample variance of observed outcome Y_i's for units receiving treatment r. We can show that $\mathfrak{s}^2(r)$ is an unbiased estimator for $\mathfrak{S}^2(r)$. Therefore, a conservative sampling variance estimator for $\hat{\tau}_{\text{tg}}(r,r')$ is

$$\widehat{\text{Var}}\{\hat{\tau}_{\text{tg}}(r,r')\} = \frac{\mathfrak{s}^2(r)}{\underline{m}_r} + \frac{\mathfrak{s}^2(r')}{\underline{m}_{r'}}.$$

Moreover, under some regularity conditions, the Wald-type confidence interval is asymptotically conservative.

Second, we show that, averaging over all possible constructions, the point estimator for $\tau_{\text{tg}}(r,r')$ is the same as $\hat{\tau}_{[a]}(r,r')$ in (13). Because the point estimator for $\tau_{\text{tg}}(r,r')$ is the standard difference-in-means for units receiving treatments r and r', it suffices to show that the average of $\underline{m}_r^{-1}\sum_{i=1}^{m} I(R_i = r)Y_i$ over all configurations of the target subpopulation is the same as $n_{[a]r}^{-1}\sum_{i:A_i=a,R_i=r} Y_i$, the average observed outcome for units with attribute a receiving treatment r. This is true because (i) a unit with attribute a receiving treatment r must be in a group with group attribute $\{a\} \cup r$, (ii) any group with group attribute $\{a\} \cup r$ has the same number of units with attribute a, and (iii) each configuration randomly picks one unit with attribute a in these groups with group attribute $\{a\} \cup r$ and calculates their average observed outcome to get $\underline{m}_r^{-1}\sum_{i=1}^{m} I(R_i = r)Y_i$.

A2. Technical details for general treatment assignments

A2.1. Lemmas

Recall that $S^2_{[a]}(r)$ and $S^2_{[a]}(r\text{-}r')$ are the finite population variances of potential outcomes $Y_i(r)$'s and individual peer effects $\tau_i(r,r')$'s among units with attribute a. We further define $S_{[a]}(r,r')$ as

the finite population covariance between the $Y_i(r)$'s and the $Y_i(r')$'s among units with attribute a, and $\widetilde{Y}_i(r) = Y_i(r) - \bar{Y}_{[A_i]}(r)$ as the centered potential outcome of unit i by subtracting the average potential outcome among units with the same attribute as unit i. We can then rewrite

$$S^2_{[a]}(r) = (n_{[a]} - 1)^{-1} \sum_{i:A_i=a} \widetilde{Y}_i^2(r), \quad S^2_{[a]}(r, r') = (n_{[a]} - 1)^{-1} \sum_{i:A_i=a} \widetilde{Y}_i(r)\widetilde{Y}_i(r'), \tag{A5}$$

and decompose the subgroup average potential outcome estimator as

$$\hat{Y}_{[a]}(r) = \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} I(R_i = r)\widetilde{Y}_i(r) + \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} I(R_i = r)\bar{Y}_{[a]}(r) \equiv B_{[a]}(r) + C_{[a]}(r),$$

and decompose the subgroup average peer effect estimator as

$$\hat{\tau}_{[a]}(r, r') = \hat{Y}_{[a]}(r) - \hat{Y}_{[a]}(r') \equiv B_{[a]}(r) + C_{[a]}(r) - B_{[a]}(r') - C_{[a]}(r'). \tag{A6}$$

The following three lemmas characterize the covariances of the terms in (A6).

Lemma A1. For $1 \leq a, a' \leq H$ and $r, r' \in \mathcal{R}$,

$$\mathrm{Cov}\{B_{[a]}(r), B_{[a']}(r')\} = \begin{cases} 0, & \text{if } a \neq a'; \\ -(n_{[a]} - 1)\frac{\pi_{[a][a]}(r,r')}{n^2_{[a]}\pi_{[a]}(r)\pi_{[a]}(r')} S_{[a]}(r, r'), & \text{if } a = a', r \neq r'; \\ (n_{[a]} - 1)\frac{\pi_{[a]}(r) - \pi_{[a][a]}(r,r)}{n^2_{[a]}\pi^2_{[a]}(r)} S^2_{[a]}(r), & \text{if } a = a', r = r'. \end{cases}$$

Lemma A2. For $1 \leq a, a' \leq H$ and $r, r' \in \mathcal{R}$, $\mathrm{Cov}\{B_{[a]}(r), C_{[a']}(r')\} = 0$.

Lemma A3. For any $1 \leq a, a' \leq H$ and $r, r' \in \mathcal{R}$,

$$\mathrm{Cov}\{C_{[a]}(r), C_{[a']}(r')\} = (n_{[a]}n_{[a']})^{-1/2} c_{[a][a']}(r, r')\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r'),$$

where $c_{[a][a']}(r, r')$ is defined in (8) in the main text.

Recall that $\overline{Y_{[a]}(r)Y_{[a]}(r')} = \{n_{[a]}(n_{[a]} - 1)\}^{-1} \sum_{i \neq j: A_i = A_j = a} Y_i(r)Y_j(r')$ is the average of the products of the potential outcomes for pairs of two different units with the same attribute a. The following lemma represents the finite population covariance $S^2_{[a]}(r, r')$ and the product of average potential outcomes $\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')$ as functions of $S^2_{[a]}(r)$, $S^2_{[a]}(r')$, $S^2_{[a]}(r\text{-}r')$ and $\overline{Y_{[a]}(r)Y_{[a]}(r')}$.

Lemma A4. For $1 \leq a \leq H$, and $r \neq r' \in \mathcal{R}$,

(a) $2S_{[a]}(r, r') = S^2_{[a]}(r) + S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r')$;

(b) $\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') = \overline{Y_{[a]}(r)Y_{[a]}(r')} + (2n_{[a]})^{-1}\{S^2_{[a]}(r) + S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r')\}$.

A2.2. Proofs of the lemmas

Proof of Lemma A1. Based on the definitions of $\pi_{[a]}(r)$ and $\pi_{[a][a']}(r,r')$, for units i and j such that $A_i = a$ and $A_j = a'$, the covariance between their treatment indicators is

$$\begin{aligned}\mathrm{Cov}\{I(R_i = r), I(R_j = r')\} &= \mathrm{pr}(R_i = r, R_j = r') - \mathrm{pr}(R_i = r)\mathrm{pr}(R_j = r') \\ &= \begin{cases} \pi_{[a][a']}(r,r') - \pi_{[a]}(r)\pi_{[a']}(r'), & \text{if } i \neq j, \\ -\pi_{[a]}(r)\pi_{[a]}(r'), & \text{if } i = j, r \neq r', \\ \pi_{[a]}(r) - \pi^2_{[a]}(r) & \text{if } i = j, r = r'. \end{cases} \end{aligned} \tag{A7}$$

Below we discuss three cases separately. (1) When $a \neq a'$, any units i and j such that $A_i = a$ and $A_j = a'$ must satisfy $i \neq j$. Therefore,

$$\begin{aligned}\mathrm{Cov}\{B_{[a]}(r), B_{[a']}(r')\} &= \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i = r)\widetilde{Y}_i(r), \{n_{[a']}\pi_{[a']}(r')\}^{-1}\sum_{j:A_j=a'} I(R_j = r')\widetilde{Y}_j(r')\right] \\ &= \{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')\}^{-1}\sum_{i:A_i=a}\sum_{j:A_j=a'}\widetilde{Y}_i(r)\widetilde{Y}_j(r')\mathrm{Cov}\{I(R_i = r), I(R_j = r')\} \\ &= \frac{\pi_{[a][a']}(r,r') - \pi_{[a]}(r)\pi_{[a']}(r')}{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')}\sum_{i:A_i=a}\sum_{j:A_j=a'}\widetilde{Y}_i(r)\widetilde{Y}_j(r'),\end{aligned}$$

where the last equality follows from (A7). We can further simplify $\mathrm{Cov}\{B_{[a]}(r), B_{[a']}(r')\}$ as

$$\mathrm{Cov}\{B_{[a]}(r), B_{[a']}(r')\} = \frac{\pi_{[a][a']}(r,r') - \pi_{[a]}(r)\pi_{[a']}(r')}{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')}\sum_{i:A_i=a}\widetilde{Y}_i(r)\sum_{j:A_j=a'}\widetilde{Y}_j(r') = 0.$$

(2) When $a = a'$ and $r \neq r'$, we need to consider the covariances between the treatment indicators of two different units and those of the same unit:

$$\begin{aligned}\mathrm{Cov}\{B_{[a]}(r), B_{[a]}(r')\} &= \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i = r)\widetilde{Y}_i(r), \{n_{[a]}\pi_{[a]}(r')\}^{-1}\sum_{j:A_j=a} I(R_j = r')\widetilde{Y}_j(r')\right] \\ &= \{n^2_{[a]}\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\sum_{i\neq j:A_i=a,A_j=a}\widetilde{Y}_i(r)\widetilde{Y}_j(r')\mathrm{Cov}\{I(R_i = r), I(R_j = r')\} \\ &\quad + \{n^2_{[a]}\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\sum_{i:A_i=a}\widetilde{Y}_i(r)\widetilde{Y}_i(r')\mathrm{Cov}\{I(R_i = r), I(R_i = r')\} \\ &= \frac{\pi_{[a][a]}(r,r') - \pi_{[a]}(r)\pi_{[a]}(r')}{n^2_{[a]}\pi_{[a]}(r)\pi_{[a]}(r')}\sum_{i\neq j:A_i=a,A_j=a}\widetilde{Y}_i(r)\widetilde{Y}_j(r') - \frac{\pi_{[a]}(r)\pi_{[a]}(r')}{n^2_{[a]}\pi_{[a]}(r)\pi_{[a]}(r')}\sum_{i:A_i=a}\widetilde{Y}_i(r)\widetilde{Y}_i(r')\end{aligned}$$

where the last equality follows from (A7). We can further simplify $\text{Cov}\{B_{[a]}(r), B_{[a]}(r')\}$ as

$$\begin{aligned}
\text{Cov}\{B_{[a]}(r), B_{[a]}(r')\} &= \frac{\pi_{[a][a]}(r,r') - \pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} \sum_{i:A_i=a} \sum_{j:A_j=a} \widetilde{Y}_i(r)\widetilde{Y}_j(r') \\
&\quad - \left(\frac{\pi_{[a][a]}(r,r') - \pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} + \frac{\pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} \right) \sum_{i:A_i=a} \widetilde{Y}_i(r)\widetilde{Y}_i(r') \\
&= 0 - \frac{(n_{[a]}-1)\pi_{[a][a]}(r,r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} (n_{[a]}-1)^{-1} \sum_{i:A_i=a} \widetilde{Y}_i(r)\widetilde{Y}_i(r') \\
&= -(n_{[a]}-1) \frac{\pi_{[a][a]}(r,r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} S_{[a]}(r,r'),
\end{aligned}$$

where the last equality follows from (A5).

(3) When $a = a'$ and $r = r'$, we similarly consider two cases with $i = j$ and $i \neq j$:

$$\begin{aligned}
\text{Var}\{B_{[a]}(r)\} &= \text{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} I(R_i = r)\widetilde{Y}_i(r), \{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{j:A_j=a} I(R_j = r)\widetilde{Y}_j(r) \right] \\
&= \{n_{[a]}^2 \pi_{[a]}^2(r)\}^{-1} \sum_{i\neq j:A_i=a,A_j=a} \widetilde{Y}_i(r)\widetilde{Y}_j(r) \text{Cov}\{I(R_i = r), I(R_j = r)\} \\
&\quad + \{n_{[a]}^2 \pi_{[a]}^2(r)\}^{-1} \sum_{i:A_i=a} \widetilde{Y}_i(r)\widetilde{Y}_i(r) \text{Cov}\{I(R_i = r), I(R_i = r)\} \\
&= \frac{\pi_{[a][a]}(r,r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \sum_{i\neq j:A_i=a,A_j=a} \widetilde{Y}_i(r)\widetilde{Y}_j(r) + \frac{\pi_{[a]}(r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \sum_{i:A_i=a} \widetilde{Y}_i^2(r)
\end{aligned}$$

where the last equality follows from (A7). We can further simplify $\text{Var}\{B_{[a]}(r)\}$ as

$$\begin{aligned}
\text{Var}\{B_{[a]}(r)\} &= \frac{\pi_{[a][a]}(r,r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \sum_{i:A_i=a} \sum_{j:A_j=a} \widetilde{Y}_i(r)\widetilde{Y}_j(r) \\
&\quad + \left(\frac{\pi_{[a]}(r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} - \frac{\pi_{[a][a]}(r,r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \right) \sum_{i:A_i=a} \widetilde{Y}_i^2(r) \\
&= 0 + \frac{\pi_{[a]}(r) - \pi_{[a][a]}(r,r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \sum_{i:A_i=a} \widetilde{Y}_i^2(r) \\
&= (n_{[a]}-1) \frac{\pi_{[a]}(r) - \pi_{[a][a]}(r,r)}{n_{[a]}^2 \pi_{[a]}^2(r)} S_{[a]}^2(r),
\end{aligned}$$

where the last equality follows from (A5). □

Proof of Lemma A2. Similarly to the proof of Lemma A1, we discuss three cases, and use (A7) to

calculate the covariance between two treatment indicators. (1) When $a \neq a'$,

$$\begin{aligned}
\mathrm{Cov}\{B_{[a]}(r), C_{[a']}(r')\} &= \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i=r)\widetilde{Y}_i(r), \{n_{[a']}\pi_{[a']}(r')\}^{-1}\sum_{j:A_j=a'} I(R_j=r')\bar{Y}_{[a']}(r')\right] \\
&= \{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')\}^{-1}\cdot\bar{Y}_{[a']}(r')\sum_{i:A_i=a}\sum_{j:A_j=a'}\widetilde{Y}_i(r)\mathrm{Cov}\{I(R_i=r), I(R_j=r')\} \\
&= \frac{\pi_{[a][a']}(r,r')-\pi_{[a]}(r)\pi_{[a']}(r')}{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')}\bar{Y}_{[a']}(r')\sum_{i:A_i=a}\sum_{j:A_j=a'}\widetilde{Y}_i(r)
\end{aligned}$$

where the last equality follows from (A7). We can further simplify $\mathrm{Cov}\{B_{[a]}(r), C_{[a']}(r')\}$ as

$$\mathrm{Cov}\{B_{[a]}(r), C_{[a']}(r')\} = \frac{\pi_{[a][a']}(r,r')-\pi_{[a]}(r)\pi_{[a']}(r')}{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')}\bar{Y}_{[a']}(r')\cdot n_{[a']}\sum_{i:A_i=a}\widetilde{Y}_i(r) = 0.$$

(2) When $a = a'$ and $r \neq r'$,

$$\begin{aligned}
\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r')\} &= \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i=r)\widetilde{Y}_i(r), \{n_{[a]}\pi_{[a]}(r')\}^{-1}\sum_{j:A_j=a} I(R_j=r')\bar{Y}_{[a]}(r')\right] \\
&= \{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\cdot\bar{Y}_{[a]}(r')\sum_{i\neq j:A_i=a,A_j=a}\widetilde{Y}_i(r)\mathrm{Cov}\{I(R_i=r), I(R_j=r')\} \\
&\quad + \{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\cdot\bar{Y}_{[a]}(r')\sum_{i:A_i=a}\widetilde{Y}_i(r)\mathrm{Cov}\{I(R_i=r), I(R_i=r')\} \\
&= \frac{\pi_{[a][a]}(r,r')-\pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')}\bar{Y}_{[a]}(r')\sum_{i\neq j:A_i=a,A_j=a}\widetilde{Y}_i(r) \\
&\quad - \frac{\pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')}\bar{Y}_{[a]}(r')\sum_{i:A_i=a}\widetilde{Y}_i(r),
\end{aligned}$$

where the last equality follows from from (A7). We can further simplify $\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r')\}$ as

$$\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r')\} = \frac{\pi_{[a][a]}(r,r')-\pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')}\bar{Y}_{[a]}(r')\cdot(n_{[a]}-1)\sum_{i:A_i=a}\widetilde{Y}_i(r) - 0 = 0.$$

(3) When $a = a'$ and $r = r'$,

$$\begin{aligned}
\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r)\} &= \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i=r)\widetilde{Y}_i(r), \{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{j:A_j=a} I(R_j=r)\bar{Y}_{[a]}(r)\right] \\
&= \{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\sum_{i\neq j:A_i=a,A_j=a}\widetilde{Y}_i(r)\bar{Y}_{[a]}(r')\mathrm{Cov}\{I(R_i=r), I(R_j=r)\} \\
&\quad + \{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\sum_{i:A_i=a}\widetilde{Y}_i(r)\bar{Y}_{[a]}(r')\mathrm{Cov}\{I(R_i=r), I(R_i=r)\}
\end{aligned}$$

$$= \frac{\pi_{[a][a]}(r,r) - \pi_{[a]}(r)\pi_{[a]}(r)}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r)} \sum_{i \neq j: A_i = a, A_j = a} \widetilde{Y}_i(r)\bar{Y}_{[a]}(r)$$
$$+ \frac{\pi_{[a]}(r) - \pi_{[a]}(r)\pi_{[a]}(r)}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r)} \sum_{i: A_i = a} \widetilde{Y}_i(r)\bar{Y}_{[a]}(r),$$

where the last equality follows from (A7). We can further simplify $\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r)\}$ as

$$\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r)\} = \frac{\pi_{[a][a]}(r,r) - \pi_{[a]}(r)\pi_{[a]}(r)}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r)} \sum_{i: A_i = a} \widetilde{Y}_i(r) \times (n_{[a]} - 1)\bar{Y}_{[a]}(r) - 0 = 0.$$

□

Proof of Lemma A3. Similary to the proofs of Lemmas A1 and A2, we discuss three cases separately, and use (A7) to calculate the covariance between two treatment indicators. (1) When $a \neq a'$,

$$\mathrm{Cov}\{C_{[a]}(r), C_{[a']}(r')\} = \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i: A_i = a} I(R_i = r)\bar{Y}_{[a]}(r), \{n_{[a']}\pi_{[a']}(r')\}^{-1} \sum_{j: A_j = a'} I(R_j = r')\bar{Y}_{[a']}(r')\right]$$
$$= \{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')\}^{-1}\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r') \sum_{i: A_i = a} \sum_{j: A_j = a'} \mathrm{Cov}\{I(R_i = r), I(R_j = r')\}$$
$$= \frac{\pi_{[a][a']}(r,r') - \pi_{[a]}(r)\pi_{[a']}(r')}{n_{[a]}n_{[a']}\pi_{[a]}(r)\pi_{[a']}(r')}\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r')n_{[a]}n_{[a']}$$
$$= \frac{\pi_{[a][a']}(r,r') - \pi_{[a]}(r)\pi_{[a']}(r')}{\pi_{[a]}(r)\pi_{[a']}(r')}\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r'),$$

where the last equality follows from (A7). Based on the definitions of $d_{[a][a']}(r,r')$ and $c_{[a][a']}(r,r')$ in (7) and (8), we can further simplify $\mathrm{Cov}\{C_{[a]}(r), C_{[a']}(r')\}$ as

$$\mathrm{Cov}\{C_{[a]}(r), C_{[a']}(r')\} = \left(\frac{\pi_{[a][a']}(r,r')}{\pi_{[a]}(r)\pi_{[a']}(r')} - 1\right)\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r') = \frac{d_{[a][a']}(r,r')}{(n_{[a]}n_{[a']})^{1/2}}\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r')$$
$$= \frac{c_{[a][a']}(r,r')}{(n_{[a]}n_{[a']})^{1/2}}\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r').$$

(2) When $a = a'$ and $r \neq r'$,

$$\mathrm{Cov}\{C_{[a]}(r), C_{[a]}(r')\} = \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i: A_i = a} I(R_i = r)\bar{Y}_{[a]}(r), \{n_{[a]}\pi_{[a]}(r')\}^{-1} \sum_{j: A_j = a} I(R_j = r')\bar{Y}_{[a]}(r')\right]$$
$$= \{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') \sum_{i \neq j: A_i = a, A_j = a} \mathrm{Cov}\{I(R_i = r), I(R_j = r')\}$$
$$+ \{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')\}^{-1}\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') \sum_{i: A_i = a} \mathrm{Cov}\{I(R_i = r), I(R_i = r')\}$$

$$= \frac{\pi_{[a][a]}(r,r') - \pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') n_{[a]}(n_{[a]}-1)$$
$$- \frac{\pi_{[a]}(r)\pi_{[a]}(r')}{n_{[a]}^2 \pi_{[a]}(r)\pi_{[a]}(r')} \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') n_{[a]},$$

where the last equality follows from (A7). Based on the definitions of $d_{[a][a]}(r,r')$ and $c_{[a][a]}(r,r')$ in (7) and (8), we can further simplify $\mathrm{Cov}\{C_{[a]}(r), C_{[a]}(r')\}$ as

$$\begin{aligned}
\mathrm{Cov}\{C_{[a]}(r), C_{[a]}(r')\} &= \left\{ \left(1 - n_{[a]}^{-1}\right) \left(\frac{\pi_{[a][a]}(r,r')}{\pi_{[a]}(r)\pi_{[a]}(r')} - 1 \right) - n_{[a]}^{-1} \right\} \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') \\
&= \left\{ \left(1 - n_{[a]}^{-1}\right) n_{[a]}^{-1} d_{[a][a]}(r,r') - n_{[a]}^{-1} \right\} \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') \\
&= \frac{\left(1 - n_{[a]}^{-1}\right) d_{[a][a]}(r,r') - 1}{n_{[a]}} \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') \\
&= \frac{c_{[a][a]}(r,r')}{n_{[a]}} \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r').
\end{aligned}$$

(3) When $a = a'$ and $r = r'$,

$$\begin{aligned}
\mathrm{Var}\{C_{[a]}(r)\} &= \mathrm{Cov}\left[\{n_{[a]}\pi_{[a]}(r)\}^{-1} \sum_{i:A_i=a} I(R_i = r)\bar{Y}_{[a]}(r), \{n_{[a']}\pi_{[a']}(r')\}^{-1} \sum_{j:A_j=a'} I(R_j = r')\bar{Y}_{[a']}(r') \right] \\
&= \{n_{[a]}^2 \pi_{[a]}^2(r)\}^{-1} \bar{Y}_{[a]}^2(r) \sum_{i \neq j: A_i = a, A_j = a} \mathrm{Cov}\{I(R_i = r), I(R_j = r)\} \\
&\quad + \{n_{[a]}^2 \pi_{[a]}^2(r)\}^{-1} \bar{Y}_{[a]}^2(r) \sum_{i:A_i=a} \mathrm{Cov}\{I(R_i = r), I(R_i = r)\} \\
&= \frac{\pi_{[a][a]}(r,r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \bar{Y}_{[a]}^2(r) n_{[a]}(n_{[a]}-1) + \frac{\pi_{[a]}(r) - \pi_{[a]}^2(r)}{n_{[a]}^2 \pi_{[a]}^2(r)} \bar{Y}_{[a]}^2(r) n_{[a]},
\end{aligned}$$

where the last equality follows from (A7). Based on the definitions of $d_{[a][a]}(r,r)$ and $c_{[a][a]}(r,r)$ in (7) and (8), we can further simplify $\mathrm{Var}\{C_{[a]}(r)\}$ as

$$\begin{aligned}
\mathrm{Var}\{C_{[a]}(r)\} &= \left\{ \left(1 - n_{[a]}^{-1}\right) \left(\frac{\pi_{[a][a]}(r,r)}{\pi_{[a]}^2(r)} - 1 \right) + n_{[a]}^{-1}\pi_{[a]}^{-1}(r) - n_{[a]}^{-1} \right\} \bar{Y}_{[a]}^2(r) \\
&= \left\{ \left(1 - n_{[a]}^{-1}\right) n_{[a]}^{-1} d_{[a][a]}(r,r) + n_{[a]}^{-1}\pi_{[a]}^{-1}(r) - n_{[a]}^{-1} \right\} \bar{Y}_{[a]}^2(r) \\
&= \frac{\left(1 - n_{[a]}^{-1}\right) d_{[a][a]}(r,r) + \pi_{[a]}^{-1}(r) - 1}{n_{[a]}} \bar{Y}_{[a]}^2(r) = \frac{c_{[a][a]}(r,r)}{n_{[a]}} \bar{Y}_{[a]}^2(r).
\end{aligned}$$

□

Proof of Lemma A4. For $1 \leq a \leq H$ and $r \neq r' \in \mathcal{R}$, by definition, we have

$$\begin{aligned} S^2_{[a]}(r) + S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r') &= (n_{[a]}-1)^{-1}\left[\sum_{i:A_i=a}\widetilde{Y}^2_i(r) + \sum_{i:A_i=a}\widetilde{Y}^2_i(r') - \sum_{i:A_i=a}\left\{\widetilde{Y}_i(r) - \widetilde{Y}_i(r')\right\}^2\right] \\ &= (n_{[a]}-1)^{-1} \times 2\sum_{i:A_i=a}\widetilde{Y}_i(r)\widetilde{Y}_i(r') = 2S_{[a]}(r,r'), \end{aligned}$$

and

$$\begin{aligned} & \overline{Y_{[a]}(r)Y_{[a]}(r')} + (2n_{[a]})^{-1}\left\{S^2_{[a]}(r) + S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r')\right\} \\ = \; & \{n_{[a]}(n_{[a]}-1)\}^{-1}\sum_{i\neq j:A_i=A_j=a} Y_i(r)Y_j(r') + n^{-1}_{[a]}S_{[a]}(r,r') \\ = \; & \{n_{[a]}(n_{[a]}-1)\}^{-1}\sum_{i\neq j:A_i=A_j=a} Y_i(r)Y_j(r') + \{n_{[a]}(n_{[a]}-1)\}^{-1}\left\{\sum_{i:A_i=a} Y_i(r)Y_i(r') - n_{[a]}\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')\right\} \\ = \; & \{n_{[a]}(n_{[a]}-1)\}^{-1}\left\{\sum_{i:A_i=a}\sum_{j:A_j=a} Y_i(r)Y_j(r') - n_{[a]}\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')\right\} \\ = \; & \{n_{[a]}(n_{[a]}-1)\}^{-1}\left\{n^2_{[a]}\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r') - n_{[a]}\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')\right\} = \bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r'). \end{aligned}$$

□

A2.3. Proofs of the theorems for general assignment mechanism

Proof of Theorem 1. First, we calculate the sampling variance of estimated subgroup average peer effect. From (A6) and Lemmas A1–A3, the covariances, $\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r)\}$, $\mathrm{Cov}\{B_{[a]}(r), C_{[a]}(r')\}$, $\mathrm{Cov}\{B_{[a]}(r'), C_{[a]}(r)\}$ and $\mathrm{Cov}\{B_{[a]}(r'), C_{[a]}(r')$, are all zero for $r \neq r'$. Therefore, the sampling variance of subgroup average peer effect estimator is

$$\begin{aligned} & \mathrm{Var}\left\{\hat{\tau}_{[a]}(r,r')\right\} \\ = \; & \mathrm{Var}\left\{B_{[a]}(r) + C_{[a]}(r) - B_{[a]}(r') - C_{[a]}(r')\right\} \\ = \; & \mathrm{Var}\left\{B_{[a]}(r)\right\} + \mathrm{Var}\left\{B_{[a]}(r')\right\} - 2\mathrm{Cov}\left\{B_{[a]}(r), B_{[a]}(r')\right\} + \mathrm{Var}\left\{C_{[a]}(r)\right\} + \mathrm{Var}\left\{C_{[a]}(r')\right\} \\ & - 2\mathrm{Cov}\left\{C_{[a]}(r), C_{[a]}(r')\right\} \\ = \; & (n_{[a]}-1)\frac{\pi_{[a]}(r) - \pi_{[a][a]}(r,r)}{n^2_{[a]}\pi^2_{[a]}(r)}S^2_{[a]}(r) + (n_{[a]}-1)\frac{\pi_{[a]}(r') - \pi_{[a][a]}(r',r')}{n^2_{[a]}\pi^2_{[a]}(r')}S^2_{[a]}(r') \\ & + 2\frac{(n_{[a]}-1)\pi_{[a][a]}(r,r')}{n^2_{[a]}\pi_{[a]}(r)\pi_{[a]}(r')}S_{[a]}(r,r') + n^{-1}_{[a]}\left\{c_{[a][a]}(r,r)\bar{Y}^2_{[a]}(r) + c_{[a][a]}(r',r')\bar{Y}^2_{[a]}(r') - 2c_{[a][a]}(r,r')\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')\right\}. \end{aligned}$$

Replacing $2S_{[a]}(r,r')$ and $\bar{Y}_{[a]}(r)\bar{Y}_{[a]}(r')$ by their expressions in Lemma A4, we can rewrite the sampling variance of $\hat{\tau}_{[a]}(r,r')$ as

$$\begin{aligned}\text{Var}\left\{\hat{\tau}_{[a]}(r,r')\right\} &= (n_{[a]}-1)\frac{\pi_{[a]}(r)-\pi_{[a][a]}(r,r)}{n_{[a]}^2\pi_{[a]}^2(r)}S_{[a]}^2(r)+(n_{[a]}-1)\frac{\pi_{[a]}(r')-\pi_{[a][a]}(r',r')}{n_{[a]}^2\pi_{[a]}^2(r')}S_{[a]}^2(r') \\ &+\frac{(n_{[a]}-1)\pi_{[a][a]}(r,r')}{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')}\left\{S_{[a]}^2(r)+S_{[a]}^2(r')-S_{[a]}^2(r\text{-}r')\right\} \\ &+n_{[a]}^{-1}\left\{c_{[a][a]}(r,r)\bar{Y}_{[a]}^2(r)+c_{[a][a]}(r',r')\bar{Y}_{[a]}^2(r')-2c_{[a][a]}(r,r')\overline{Y_{[a]}(r)Y_{[a]}(r')}\right\} \\ &-\frac{2c_{[a][a]}(r,r')}{n_{[a]}}\frac{S_{[a]}^2(r)+S_{[a]}^2(r')-S_{[a]}^2(r\text{-}r')}{2n_{[a]}}. \end{aligned} \tag{A8}$$

We then combine the terms and calculate the cofficients of $S_{[a]}^2(r), S_{[a]}^2(r')$ and $S_{[a]}^2(r\text{-}r')$ in (A8), separately. From the definitions of $c_{[a][a]}(r,r')$ and $b_{[a]}(r)$, we can simplify the coefficient of $S_{[a]}^2(r)$ as

$$n_{[a]}^{-1}\left\{(1-n_{[a]}^{-1})\left(\pi_{[a]}^{-1}(r)-\frac{\pi_{[a][a]}(r,r)}{\pi_{[a]}^2(r)}+\frac{\pi_{[a][a]}(r,r')}{\pi_{[a]}(r)\pi_{[a]}(r')}\right)-n_{[a]}^{-1}c_{[a][a]}(r,r')\right\}=n_{[a]}^{-1}b_{[a]}(r),$$

and similarly simplify the coefficient of $S_{[a]}^2(r')$ as $n_{[a]}^{-1}b_{[a]}(r')$. We can also simplify the coefficient of $S_{[a]}^2(r\text{-}r')$ as

$$-\frac{(n_{[a]}-1)\pi_{[a][a]}(r,r')}{n_{[a]}^2\pi_{[a]}(r)\pi_{[a]}(r')}+\frac{c_{[a][a]}(r,r')}{n_{[a]}^2}=-n_{[a]}^{-1}.$$

Therefore, (A8) reduces to

$$\begin{aligned}\text{Var}\left\{\hat{\tau}_{[a]}(r,r')\right\} &= n_{[a]}^{-1}\left\{b_{[a]}(r)S_{[a]}^2(r)+b_{[a]}(r')S_{[a]}^2(r')-S_{[a]}^2(r\text{-}r')\right\} \\ &+n_{[a]}^{-1}\left\{c_{[a][a]}(r,r)\bar{Y}_{[a]}^2(r)+c_{[a][a]}(r',r')\bar{Y}_{[a]}^2(r')-2c_{[a][a]}(r,r')\overline{Y_{[a]}(r)Y_{[a]}(r')}\right\}.\end{aligned}$$

Second, we calculate the covariance between two estimated subgroup average peer effects. For $a \neq a'$ and $r \neq r' \in \mathcal{R}$, according to Lemmas A1–A3, the covariances, $\text{Cov}\{B_{[a]}(r), B_{[a']}(r)\}$, $\text{Cov}\{B_{[a]}(r), B_{[a']}(r')\}$, $\text{Cov}\{B_{[a]}(r'), B_{[a']}(r)\}$, $\text{Cov}\{B_{[a]}(r'), B_{[a']}(r')\}$, $\text{Cov}\{B_{[a]}(r), C_{[a']}(r)\}$, $\text{Cov}\{B_{[a]}(r), C_{[a']}(r')\}$, $\text{Cov}\{B_{[a]}(r'), C_{[a']}(r)\}$ and $\text{Cov}\{B_{[a]}(r'), C_{[a']}(r')\}$, are all zero. Therefore the sampling covariance between $\hat{\tau}_{[a]}(r,r')$ and $\hat{\tau}_{[a']}(r,r')$ is

$$\begin{aligned}&\text{Cov}\left\{\hat{\tau}_{[a]}(r,r'),\hat{\tau}_{[a']}(r,r')\right\} \\ =\ &\text{Cov}\left\{B_{[a]}(r)+C_{[a]}(r)-B_{[a]}(r')-C_{[a]}(r'),B_{[a']}(r)+C_{[a']}(r)-B_{[a']}(r')-C_{[a']}(r')\right\} \\ =\ &\text{Cov}\left\{C_{[a]}(r),C_{[a']}(r)\right\}+\text{Cov}\left\{C_{[a]}(r'),C_{[a']}(r')\right\}-\text{Cov}\left\{C_{[a]}(r),C_{[a']}(r')\right\}-\text{Cov}\left\{C_{[a]}(r'),C_{[a']}(r)\right\} \\ =\ &(n_{[a]}n_{[a']})^{-1/2}\left\{c_{[a][a']}(r,r)\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r)+c_{[a][a']}(r',r')\bar{Y}_{[a]}(r')\bar{Y}_{[a']}(r')\right.\end{aligned}$$

$$-c_{[a][a']}(r,r')\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r') - c_{[a][a']}(r',r)\bar{Y}_{[a]}(r')\bar{Y}_{[a']}(r)\Big\},$$

where the last equality follows from Lemma A3.

Third, we calculate the variance of the average peer effect estimator:

$$\text{Var}\left\{\hat{\tau}(r,r')\right\} = \text{Var}\left\{\sum_{a=1}^{H} w_{[a]}\hat{\tau}_{[a]}(r,r')\right\} = \sum_{a=1}^{H} w_{[a]}^2 \text{Var}\left\{\hat{\tau}_{[a]}(r,r')\right\} + \sum_{a=1}^{H}\sum_{a'\neq a} w_{[a]}w_{[a']}\text{Cov}\left\{\hat{\tau}_{[a]}(r,r'),\hat{\tau}_{[a']}(r,r')\right\}.$$

Using the variances and covariances between the subgroup average peer effect estimators, we can simplify the variance of $\hat{\tau}(r,r')$ as

$$\begin{aligned}
\text{Var}\left\{\hat{\tau}(r,r')\right\} &= n^{-1}\sum_{a=1}^{H} w_{[a]}\left\{b_{[a]}(r)S^2_{[a]}(r) + b_{[a]}(r')S^2_{[a]}(r') - S^2_{[a]}(r\text{-}r')\right\} \\
&\quad + n^{-1}\sum_{a=1}^{H} w_{[a]}\left\{c_{[a][a]}(r,r)\bar{Y}^2_{[a]}(r) + c_{[a][a]}(r',r')\bar{Y}^2_{[a]}(r') - 2c_{[a][a]}(r,r')\overline{Y_{[a]}(r)Y_{[a]}(r')}\right\} \\
&\quad + n^{-1}\sum_{a=1}^{H}\sum_{a'\neq a}(w_{[a]}w_{[a']})^{1/2}\left\{c_{[a][a']}(r,r)\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r) + c_{[a][a']}(r',r')\bar{Y}_{[a]}(r')\bar{Y}_{[a']}(r')\right. \\
&\quad \left. - c_{[a][a']}(r,r')\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r') - c_{[a][a']}(r',r)\bar{Y}_{[a]}(r')\bar{Y}_{[a']}(r)\right\}.
\end{aligned}$$

□

Proof of Theorem 2. First, we prove that $s^2_{[a]}(r)$ defined in Section 3.3 is unbiased for the finite population variance $S^2_{[a]}(r)$. Note that

$$\begin{aligned}
E\left\{\hat{Y}^2_{[a]}(r)\right\} &= E\left\{\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i=r)Y_i(r) \times \{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{j:A_j=a} I(R_j=r)Y_j(r)\right\} \\
&= \{n^2_{[a]}\pi^2_{[a]}(r)\}^{-1}\sum_{i\neq j:A_i=A_j=a} Y_i(r)Y_j(r)\text{pr}(R_i=r,R_j=r) \\
&\quad + \{n^2_{[a]}\pi^2_{[a]}(r)\}^{-1}\sum_{i:A_i=a} Y_i^2(r)\text{pr}(R_i=r,R_i=r) \\
&= \frac{\pi_{[a][a]}(r,r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i\neq j:A_i=A_j=a} Y_i(r)Y_j(r) + \frac{\pi_{[a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a} Y_i^2(r),
\end{aligned}$$

where the last equality follows from the definitions of $\pi_{[a]}(r)$ and $\pi_{[a][a]}(r,r)$. We can then simplify $E\{\hat{Y}^2_{[a]}(r)\}$ as

$$\begin{aligned}
E\left\{\hat{Y}^2_{[a]}(r)\right\} &= \frac{\pi_{[a][a]}(r,r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a}\sum_{j:A_j=a} Y_i(r)Y_j(r) + \frac{\pi_{[a]}(r)-\pi_{[a][a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a} Y_i^2(r) \\
&= \frac{\pi_{[a][a]}(r,r)}{\pi^2_{[a]}(r)}\bar{Y}^2_{[a]}(r) + \frac{\pi_{[a]}(r)-\pi_{[a][a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a} Y_i^2(r). \qquad \text{(A9)}
\end{aligned}$$

The mean of $s^2_{[a]}(r)$ is

$$\begin{aligned}
E\left\{s^2_{[a]}(r)\right\} &= \frac{n_{[a]}\pi^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r)}\left[\frac{n_{[a]}+c_{[a][a]}(r,r)}{n^2_{[a]}\pi_{[a]}(r)}\sum_{i:A_i=a}E\left\{I(R_i=r)\right\}Y^2_i(r)-E\left\{\hat{Y}^2_{[a]}(r)\right\}\right] \\
&= \frac{n_{[a]}\pi^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r)}\left[\frac{n_{[a]}+c_{[a][a]}(r,r)}{n^2_{[a]}}\sum_{i:A_i=a}Y^2_i(r)-E\left\{\hat{Y}^2_{[a]}(r)\right\}\right] \\
&= \frac{n_{[a]}\pi^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r)}\left[\frac{(n_{[a]}-1)\pi_{[a][a]}(r,r)+\pi_{[a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a}Y^2_i(r)-E\left\{\hat{Y}^2_{[a]}(r)\right\}\right],
\end{aligned}$$

where the last equality follows from the definition of $c_{[a][a]}(r,r)$ in (8). Using (A9), we can further simplify $E\{s^2_{[a]}(r)\}$ as

$$\begin{aligned}
E\left\{s^2_{[a]}(r)\right\} &= \frac{n_{[a]}\pi^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r)}\left\{\frac{(n_{[a]}-1)\pi_{[a][a]}(r,r)+\pi_{[a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a}Y^2_i(r)\right. \\
&\qquad \left.-\frac{\pi_{[a][a]}(r,r)}{\pi^2_{[a]}(r)}\bar{Y}^2_{[a]}(r)-\frac{\pi_{[a]}(r)-\pi_{[a][a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a}Y^2_i(r)\right\} \\
&= \frac{n_{[a]}\pi^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r)}\left\{\frac{\pi_{[a][a]}(r,r)}{n_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a}Y^2_i(r)-\frac{\pi_{[a][a]}(r,r)}{\pi^2_{[a]}(r)}\bar{Y}^2_{[a]}(r)\right\} \\
&= (n_{[a]}-1)^{-1}\left\{\sum_{i:A_i=a}Y^2_i(r)-n_{[a]}\bar{Y}^2_{[a]}(r)\right\}=S^2_{[a]}(r).
\end{aligned}$$

Second, we prove the unbiasedness of the estimator for $\bar{Y}^2_{[a]}(r)$. For $1\le a\le H$ and $r\in\mathcal{R}$, accoring to (A9) and the unbiasedness of $s^2_{[a]}(r)$ for $S^2_{[a]}(r)$,

$$\begin{aligned}
&E\left[\frac{n_{[a]}\hat{Y}_{[a]}(r)^2-\{b_{[a]}(r)-1\}s^2_{[a]}(r)}{n_{[a]}+c_{[a][a]}(r,r)}\right] \\
=\ &\frac{n_{[a]}E\{\hat{Y}_{[a]}(r)^2\}-\{b_{[a]}(r)-1\}E\{s^2_{[a]}(r)\}}{n_{[a]}+c_{[a][a]}(r,r)} \\
=\ &\frac{n_{[a]}}{n_{[a]}+c_{[a][a]}(r,r)}\left\{\frac{\pi_{[a][a]}(r,r)}{\pi^2_{[a]}(r)}\bar{Y}^2_{[a]}(r)+\frac{\pi_{[a]}(r)-\pi_{[a][a]}(r)}{n^2_{[a]}\pi^2_{[a]}(r)}\sum_{i:A_i=a}Y^2_i(r)\right\}-\frac{b_{[a]}(r)-1}{n_{[a]}+c_{[a][a]}(r,r)}S^2_{[a]}(r),
\end{aligned}$$

which, based on the definitions of $c_{[a][a]}(r,r)$ and $b_{[a]}(r)$, further reduces to

$$\begin{aligned}
&\frac{n_{[a]}}{(n_{[a]}-1)\pi_{[a][a]}(r,r)+\pi_{[a]}(r)}\left\{\pi_{[a][a]}(r,r)\bar{Y}^2_{[a]}(r)+\frac{\pi_{[a]}(r)-\pi_{[a][a]}(r)}{n^2_{[a]}}\sum_{i:A_i=a}Y^2_i(r)\right\} \\
&-\frac{\{\pi_{[a]}(r)-\pi_{[a][a]}(r,r)\}}{n_{[a]}\{(n_{[a]}-1)\pi_{[a][a]}(r,r)+\pi_{[a]}(r)\}}\left\{\sum_{i:A_i=a}Y^2_i(r)-n_{[a]}\bar{Y}^2_{[a]}(r)\right\}
\end{aligned}$$

$$= \frac{n_{[a]}\pi_{[a][a]}(r,r)\bar{Y}^2_{[a]}(r) + \{\pi_{[a]}(r) - \pi_{[a][a]}(r,r)\}\bar{Y}^2_{[a]}(r)}{(n_{[a]}-1)\pi_{[a][a]}(r,r) + \pi_{[a]}(r)} = \bar{Y}^2_{[a]}(r).$$

Third, we prove the unbiasedness of the estimator for $\overline{Y_{[a]}(r)Y_{[a]}(r')}$. For $1 \le a \le H$ and $r \ne r' \in \mathcal{R}$,

$$\begin{aligned}
& E\left\{\frac{n_{[a]}}{n_{[a]}-1}\frac{\pi_{[a]}(r)\pi_{[a]}(r')}{\pi_{[a][a]}(r,r')}\hat{Y}_{[a]}(r)\hat{Y}_{[a]}(r')\right\} \\
&= \frac{n_{[a]}}{n_{[a]}-1}\frac{\pi_{[a]}(r)\pi_{[a]}(r')}{\pi_{[a][a]}(r,r')}E\left\{\{n_{[a]}\pi_{[a]}(r)\}^{-1}\sum_{i:A_i=a} I(R_i=r)Y_i(r) \times \{n_{[a]}\pi_{[a]}(r')\}^{-1}\sum_{j:A_j=a} I(R_j=r')Y_j(r')\right\} \\
&= \frac{n_{[a]}}{n_{[a]}-1}\frac{1}{n^2_{[a]}\pi_{[a][a]}(r,r')}\sum_{i,j:A_i=A_j=a} Y_i(r)Y_j(r')\mathrm{pr}(R_i=r,R_j=r') \\
&= \frac{1}{n_{[a]}(n_{[a]}-1)\pi_{[a][a]}(r,r')}\left\{\sum_{i\ne j:A_i=A_j=a} Y_i(r)Y_j(r')\mathrm{pr}(R_i=r,R_j=r') + \sum_{i:A_i=a} Y_i(r)Y_i(r')\mathrm{pr}(R_i=r,R_i=r')\right\} \\
&= \frac{1}{n_{[a]}(n_{[a]}-1)\pi_{[a][a]}(r,r')}\left\{\pi_{[a][a]}(r,r')\sum_{i\ne j:A_i=A_j=a} Y_i(r)Y_j(r') + 0\right\} = \overline{Y_{[a]}(r)Y_{[a]}(r')},
\end{aligned}$$

where the second last equality holds because $\mathrm{pr}(R_i=r,R_i=r')=0$ for $r\ne r'\in\mathcal{R}$.

Fourth, we prove that, for $a\ne a'$ and $r,r'\in\mathcal{R}$, $\pi_{[a]}(r)\pi_{[a']}(r')/\pi_{[a][a']}(r,r')\cdot\hat{Y}_{[a]}(r)\hat{Y}_{[a']}(r')$ is unbiased for $\bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r')$. For $a\ne a'$ and $r,r'\in\mathcal{R}$, any units (i,j) such that $A_i=a$ and $A_j=a'$ must satisfy $i\ne j$, and therefore

$$\begin{aligned}
& E\left\{\frac{\pi_{[a]}(r)\pi_{[a']}(r')}{\pi_{[a][a']}(r,r')}\hat{Y}_{[a]}(r)\hat{Y}_{[a']}(r')\right\} \\
=\ & \frac{\pi_{[a]}(r)\pi_{[a']}(r')}{\pi_{[a][a']}(r,r')}E\left\{n^{-1}_{[a]}\sum_{i:A_i=a}\pi^{-1}_{[a]}(r)I(R_i=r)Y_i(r) \times n^{-1}_{[a']}\sum_{j:A_j=a'}\pi^{-1}_{a'}(r')I(R_j=r')Y_j(r')\right\} \\
=\ & \{n_{[a]}n_{[a']}\pi_{[a][a']}(r,r')\}^{-1}\sum_{i:A_i=a}\sum_{j:A_j=a'} Y_i(r)Y_j(r')\mathrm{pr}(R_i=r,R_j=r') \\
=\ & (n_{[a]}n_{[a']})^{-1}\sum_{i:A_i=a}\sum_{j:A_j=a'} Y_i(r)Y_j(r') = \bar{Y}_{[a]}(r)\bar{Y}_{[a']}(r').
\end{aligned}$$

□

A3. More technical details about complete randomization

Proof of Proposition 2. We first show that the numerical implementation in Section 2.4.2 generates treatment assignments under complete randomization. For any treatment assignment z with $L(z)=z$, by definition, there are l_t groups with group attribute set g_t for $1\le t\le T$. Thus, there

are $\prod_{t=1}^{T} l_t!$ ways to arrange these m groups such that the first l_1 groups have group attribute set g_1, the next l_2 groups have group attribute g_2, ..., and the last l_T groups have group attribute g_T. Each of the $\prod_{t=1}^{T} l_t!$ arrangements is a group assignment from the numerical implementation, and all group assignments from the numerical implementation have the same probability. Therefore, under the numerical implementation, any assignment z with $L(z) = z$ corresponds to $\prod_{t=1}^{T} l_t!$ realizations and will have the same probability.

We then prove Proposition 2. From the above discussion, it is equivalent to consider the distribution of $(R_1, \ldots, R_n)$ under the group assignment generated from the numerical implementation. Under the numerical implementation, for each $1 \leq a \leq H$ and the $n_{[a]}$ units with attribute a, the $l_1 \times g_1(a)$ units in the first l_1 groups must receive treatment $g_1 \setminus \{a\}$, the next $l_2 \times g_2(a)$ units in the next l_2 groups must receive treatment $g_2 \setminus \{a\}$, ..., and the last $l_T \times g_T(a)$ units in the last l_T groups must receive treatment $g_T \setminus \{a\}$, and the assignments of the $n_{[a]}$ units into these m groups have the same probability. From the relationship between $n_{[a]r}$ and $L_t(z)g_t(a)$ in (4), the first conclusion (1) in Proposition 2 holds. The second conclusion (2) in Proposition 2 follows directly from the independence among the group assignments for units with different attributes under the numerical implementation. □

As a direct consequence of Proposition 2, we have the following results characterizing the probability law of complete randomization, and we will use them in later proofs.

Proposition A2. Under Assumptions 1 and 2, and under complete randomization defined in Section 2.4.2, for $1 \leq a, a' \leq H$ and $r, r' \in \mathcal{R}$, we have $\pi_{[a]}(r) = n_{[a]r}/n_{[a]}$, $b_{[a]}(r) = n_{[a]}/n_{[a]r}$, $c_{[a][a']}(r, r') = 0$, and

$$\pi_{[a][a']}(r,r') = \begin{cases} \frac{n_{[a]r} n_{[a']r'}}{n_{[a]} n_{[a']}}; \\ \frac{n_{[a]r} n_{[a]r'}}{n_{[a]}(n_{[a]}-1)}; \\ \frac{n_{[a]r}(n_{[a]r}-1)}{n_{[a]}(n_{[a]}-1)}; \end{cases} \qquad d_{[a][a']}(r,r') = \begin{cases} 0, & \text{if } a \neq a'; \\ \frac{n_{[a]}}{n_{[a]}-1}, & \text{if } a = a', r \neq r'; \\ -\frac{n_{[a]}(n_{[a]}-n_{[a]r})}{(n_{[a]}-1)n_{[a]r}}, & \text{if } a = a', r = r'. \end{cases}$$

The formulas of $\pi_{[a]}(r)$ and $\pi_{[a][a]}(r, r')$ are standard in completely randomized experiments with multiple treatments, the formula of $\pi_{[a][a']}(r, r')$ with $a \neq a'$ follows from the independence between treatments of units with different attributes, and the formulas of $d_{[a][a']}(r, r')$, $c_{[a][a']}(r, r')$ and $b_{[a]}(r)$ follow from their definitions in (7)–(9).

Proof of Theorem 3. We first consider the point and interval estimator for the subgroup average peer effect. Under complete randomization, for the $n_{[a]}$ units with attribute a, we are essentially conducting a complete randomized experiments with $n_{[a]r}$ units receiving treatment r. Based on Lemma A4 and the regularity condition (ii) in Condition 1, the finite population covariance between potential outcomes $S_{[a]}(r, r')$ has a limit. From Li and Ding (2017, Theorem 5), $\hat{\tau}_{[a]}(r, r')$ is asymptotically Normal:

$$\sqrt{n_{[a]}} \left\{ \hat{\tau}_{[a]}(r,r') - \tau_{[a]}(r,r') \right\} \xrightarrow{d} \mathcal{N}\left(0, \lim_{n\to\infty} n_{[a]} \mathrm{Var}\{\hat{\tau}_{[a]}(r,r')\}\right),$$

where $\lim_{n\to\infty} n_{[a]}\text{Var}\{\hat{\tau}_{[a]}(r,r')\}$ exists due to the convergence of proportions of units receiving different treatments $n_{[a]r}/n_{[a]}$ and the finite population variances of potential outcomes and individual peer effects $S^2_{[a]}(r)$ and $S^2_{[a]}(r\text{-}r')$.

Moreover, according to Li and Ding (2017, Proposition 3), the sample variance of observed outcomes in the subgroup consisting of units with attribute a receiving treatment r, $s^2_{[a]}(r)$, is consistent for the population analogue $S^2_{[a]}(r)$, in the sense that $s^2_{[a]}(r) - S^2_{[a]}(r) \xrightarrow{p} 0$. Thus, the variance estimator $\hat{V}_{[a]}(r,r')$ satisfies

$$n_{[a]}\hat{V}_{[a]}(r,r') - n_{[a]}\text{Var}\{\hat{\tau}_{[a]}(r,r')\} - S^2_{[a]}(r\text{-}r') = \frac{n_{[a]}}{n_{[a]r}}\left\{s^2_{[a]}(r) - S^2_{[a]}(r)\right\} + \frac{n_{[a]}}{n_{[a]r'}}\left\{s^2_{[a]}(r') - S^2_{[a]}(r')\right\} \xrightarrow{p} 0.$$

Therefore, the Wald-type confidence interval for $\tau_{[a]}(r,r')$ is asymptotically conservative.

Second, we consider the confidence interval for the average peer effect $\tau(r,r')$. Based on Slutsky's theorem, $\hat{\tau}(r,r')$ is asymptotically Normal:

$$\sqrt{n}\left\{\hat{\tau}(r,r') - \tau(r,r')\right\} = \sum_{a=1}^{H}\sqrt{w_{[a]}}\sqrt{n_{[a]}}\{\hat{\tau}_{[a]}(r,r') - \tau_{[a]}(r,r')\} \xrightarrow{d} \mathcal{N}\left(0,\ \lim_{n\to\infty} n\text{Var}\{\hat{\tau}(r,r')\}\right).$$

Moreover, the variance estimator $\hat{V}(r,r')$ satisfies that

$$n\hat{V}(r,r') - n\text{Var}\{\hat{\tau}(r,r')\} - \sum_{a=1}^{H} w_{[a]}S^2_{[a]}(r\text{-}r') = \sum_{a=1}^{H} w_{[a]}\left\{n_{[a]}\hat{V}_{[a]}(r,r') - n_{[a]}\text{Var}\{\hat{\tau}_{[a]}(r,r')\} - S^2_{[a]}(r\text{-}r')\right\} \xrightarrow{p} 0.$$

Therefore, the Wald-type confidence interval for $\tau(r,r')$ is asymptotically conservative. □

Proof of Theorem 4. We prove the three conclusions in Theorem 4 as follows.

First, let $|\mathcal{R}|$ dimensional column vectors

$$Y_i(\mathcal{R}) = (Y_i(r_1),\ldots,Y_i(r_{\mathcal{R}}))^\top,\quad \bar{Y}_{[a]}(\mathcal{R}) = (\bar{Y}_{[a]}(r_1),\ldots,\bar{Y}_{[a]}(r_{|\mathcal{R}|}))^\top,\quad \hat{Y}_{[a]}(\mathcal{R}) = (\hat{Y}_{[a]}(r_1),\ldots,\hat{Y}_{[a]}(r_{|\mathcal{R}|}))^\top$$

consist of unit i's all potential outcomes, all subgroup average potential outcomes, and all subgroup average potential outcome estimators, respectively. Then

$$\theta_i(\mathcal{R}) = \Gamma Y_i(\mathcal{R}),\quad \theta_{[a]}(\mathcal{R}) = \Gamma\bar{Y}_{[a]}(\mathcal{R}),\quad \hat{\theta}_{[a]}(\mathcal{R}) = \Gamma\hat{Y}_{[a]}(\mathcal{R}).$$

Based on the equivalence relationship in Proposition 2 and the variance formula in Li and Ding (2017, Theorem 3), the sampling covariance matrix of $\hat{Y}_{[a]}(\mathcal{R})$ under complete randomization is

$$\text{Cov}\{\hat{Y}_{[a]}(\mathcal{R})\} = \text{diag}\left\{\frac{S^2_{[a]}(r_1)}{n_{[a]r_1}},\ldots,\frac{S^2_{[a]}(r_{|\mathcal{R}|})}{n_{[a]r_{|\mathcal{R}|}}}\right\} - \frac{1}{n_{[a]}(n_{[a]}-1)}\sum_{i:A_i=a}\{Y_i(\mathcal{R}) - \bar{Y}_{[a]}(\mathcal{R})\}\{Y_i(\mathcal{R}) - \bar{Y}_{[a]}(\mathcal{R})\}^\top,$$

which implies the sampling covariance of $\hat{\theta}_{[a]}(\mathcal{R}) = \Gamma\hat{Y}_{[a]}(\mathcal{R})$.

Second, the regularity conditions of Theorem 4 and Li and Ding (2017, Theorem 5) imme-

diately imply the asymptotic Normality of $\sqrt{n_{[a]}}\{\hat{Y}_{[a]}(\mathcal{R}) - \bar{Y}_{[a]}(\mathcal{R})\}$, which further implies the asymptotic Normality of $\sqrt{n_{[a]}}\{\hat{\theta}_{[a]}(\mathcal{R}) - \theta_{[a]}(\mathcal{R})\}$.

Third, according to Li and Ding (2017, Proposition 3), $s^2_{[a]}(r)$ is consistent for $S^2_{[a]}(r)$. Moreover, the second term in the covariance formula of $\hat{\theta}_{[a]}(\mathcal{R})$ is a positive semi-definite matrix. Therefore, the Wald-type confidence set using variance estimator (17) is asymptotically conservative.

Note that the second term in the covariance formula of $\hat{\theta}_{[a]}(\mathcal{R})$ is actually the finite population covariance matrix of $(\theta_i(r_1), \theta_i(r_2), \ldots, \theta_i(r_{|\mathcal{R}|}))^\top$ for units with attribute a scaled by $n_{[a]}^{-1}$, and these centered individual potential outcomes $\theta_i(r)$'s can be represented as linear functions of the individual peer effects $\tau_i(r, r')$'s. Thus, when the individual peer effects for units with the same attribute are additive, the finite population covariance matrix of $(\theta_i(r_1), \theta_i(r_2), \ldots, \theta_i(r_{|\mathcal{R}|}))^\top$ for units with attribute a are zero, and the Wald-type confidence sets for $\theta_{[a]}(\mathcal{R})$ become asymptotically exact. □

A4. More on random partitioning

In this section, we discuss details of random partitioning. In particular, we give the formulas for $\pi_{[a]}(r)$ and $\pi_{[a][a']}(r, r')$, based on which we can get the formulas for $d_{[a][a']}(r, r')$, $c_{[a][a']}(r, r')$ and $b_{[a]}(r)$, the unbiased point estimators for peer effects, the sampling variances of peer effects estimators, and the corresponding variance estimators.

Each $r \in \mathcal{R}$ is a set containing K unordered but replicable elements from $\{1, 2, \ldots, H\}$. Let $r(a)$ be the number of elements in set r that are equal to a. If a itself belongs to r, let $r \setminus \{a\}$ be the set containing the remaining $K-1$ elements, by deleting an element a from the set r.

Theorem A1. Under random partitioning, for $1 \leq a, a' \leq H$ and $r, r' \in \mathcal{R}$, the probability that a unit i with attribute $A_i = a$ receives treatment r is

$$\pi_{[a]}(r) = \text{pr}(R_i = r) = \frac{\binom{n_{[a]}-1}{r(a)} \prod_{1 \leq q \leq H, q \neq a} \binom{n_{[q]}}{r(q)}}{\binom{n-1}{K}}, \tag{A10}$$

and the probability that two different units ($i \neq j$) with attributes $A_i = a$ and $A_j = a'$ receive treatments r and r' is

$$\pi_{[a][a']}(r, r') = \text{pr}(R_i = r, R_j = r') = \frac{K}{n-1} \cdot \psi_{[a][a']}(r, r') + \frac{n-K-1}{n-1} \cdot \phi_{[a][a']}(r, r'), \tag{A11}$$

where

$$\psi_{[a][a']}(r, r') = \begin{cases} \frac{\binom{n_{[a]}-1}{r(a)} \binom{n_{[a']}-1}{r(a')-1} \prod_{1 \leq q \leq H, q \neq a, a'} \binom{n_{[q]}}{r(q)}}{\binom{n-2}{K-1}}, & \text{if } a' \in r, a \in r', r \setminus \{a'\} = r' \setminus \{a\}, a \neq a', \\ \frac{\binom{n_{[a]}-2}{r(a)-1} \prod_{1 \leq q \leq H, q \neq a} \binom{n_{[q]}}{r(q)}}{\binom{n-2}{K-1}}, & \text{if } a' \in r, a \in r', r \setminus \{a'\} = r' \setminus \{a\}, a = a', \\ 0, & \text{otherwise,} \end{cases} \tag{A12}$$

and

$$\phi_{[a][a']}(r,r') = \begin{cases} \dfrac{\binom{n_{[a]}-1}{r(a)}\binom{n_{[a]}-1-r(a)}{r'(a)}\binom{n_{[a']}-1}{r(a')}\binom{n_{[a']}-1-r(a')}{r'(a')}\prod_{1\le q\le H, q\neq a,a'}\left\{\binom{n_{[q]}}{r(q)}\binom{n_{[q]}-r(q)}{r'(q)}\right\}}{\binom{n-2}{K}\binom{n-2-K}{K}}, & \text{if } a \neq a', \\ \dfrac{\binom{n_{[a]}-2}{r(a)}\binom{n_{[a]}-2-r(a)}{r'(a)}\prod_{1\le q\le H, q\neq a}\left\{\binom{n_{[q]}}{r(q)}\binom{n_{[q]}-r(q)}{r'(q)}\right\}}{\binom{n-2}{K}\binom{n-2-K}{K}}, & \text{if } a = a'. \end{cases} \tag{A13}$$

We give some intuition to explain the formulas of $\pi_{[a]}(r)$ and $\pi_{[a][a']}(r,r')$ under random partitioning. First, in (A10), the denominator $\binom{n-1}{K}$ denotes the total number of possible peers for unit i, and the numerator denotes the number of possible peers such that unit i receives treatment r. Second, for any two different units i and j with attributes a and a', we consider two cases according to whether units i and j are in the same group or not. The coefficients $K/(n-1)$ and $(n-K-1)/(n-1)$ in (A11) are the probabilities that units i and j are in the same group and not in the same group, respectively. Correspondingly, $\psi_{[a][a']}(r,r')$ and $\phi_{[a][a']}(r,r')$ represent the conditional probabilities that units i and j receive treatments r and r' given that i and j are and are not in the same group.

When units i and j are in the same group, they have $K-1$ common peers, and therefore, the treatment R_i of unit i consists of unit j's attribute and the $K-1$ common peers' attributes, and the treatment R_j consists of unit i's attribute and the $K-1$ common peers' attributes. Therefore, $\psi_{[a][a']}(r,r')$ is positive if and only if $a' \in r, a \in r'$ and $r \setminus \{a'\} = r' \setminus \{a\}$. In (A12), when $\psi_{[a][a']}(r,r') \neq 0$, the denominator $\binom{n-2}{K-1}$ counts the number of possible $K-1$ units in the same group as units i and j, and the numerator counts the number of possible $K-1$ units in the same group as units i and j such that units i and j receive treatments r and r'. In (A13), the denominator $\binom{n-2}{K}\binom{n-2-K}{K}$ counts the number of possible peers for units i and j, and the numerator counts the number of possible peers for units i and j such that units i and j receive treatments r and r'.

Proof of Theorem A1. First, we calculate $\pi_{[a]}(r)$. Assume that unit i has attribute $A_i = a$. The total number of possible peers of unit i is $\binom{n-1}{K}$, and the total number of possible peers of unit i such that unit i receives treatment r is

$$\binom{n_{[a]}-1}{r(a)} \prod_{1\le q\le H, q\neq a} \binom{n_{[q]}}{r(q)},$$

where $\binom{n_{[a]}-1}{r(a)}$ counts the number of possible choice of the peers of unit i with attribute a, and $\binom{n_{[q]}}{r(q)}$ counts the number of possible choice of the peers of unit i with attribute $q \neq a$. Under random partitioning, any other K units have the same probability to be in the same group as unit i. Therefore, (A10) holds.

Second, we calculate $\pi_{[a][a']}(r,r')$. Assume that $i \neq j$ are two units with attributes $A_i = a$ and $A_j = a'$. Under random partitioning, we can decompose the probability $\text{pr}(R_i = r, R_j = r')$ into two parts according to whether units i and j are in the same group:

$$\pi_{[a][a']}(r,r') = \text{pr}(R_i = r, R_j = r') = \text{pr}(j \in Z_i, R_i = r, R_j = r') + \text{pr}(j \notin Z_i, R_i = r, R_j = r')$$

$$= \text{pr}(j \in Z_i)\text{pr}(R_i = r, R_j = r' \mid j \in Z_i) + \text{pr}(j \notin Z_i)\text{pr}(R_i = r, R_j = r' \mid j \notin Z_i). \tag{A14}$$

Because any other K units have the same probability to be the peers of unit i, by symmetry, $\text{pr}(j \in Z_i) = K/(n-1)$ and $\text{pr}(j \notin Z_i) = 1 - K/(n-1)$. We then consider the two conditional probabilities $\psi_{[a][a']}(r, r') \equiv \text{pr}(R_i = r, R_j = r' \mid j \in Z_i)$, and $\phi_{[a][a']}(r, r') \equiv \text{pr}(R_i = r, R_j = r' \mid j \notin Z_i)$.

When units i and j are in the same group, units i and j are peers of each other and they have $K-1$ common peers. Therefore, they have positive probability to receive treatments r and r' if and only if r and r' satisfy $a' \in r$, $a \in r'$, and $r \setminus \{a'\} = r' \setminus \{a\}$. When $a' \in r$, $a \in r'$, and $r \setminus \{a'\} = r' \setminus \{a\}$, the total number of possible peers of units i and j such that units i and j receive treatments r and r' is

$$\begin{cases} \binom{n_{[a]}-1}{r(a)}\binom{n_{[a']}-1}{r(a')-1}\prod_{1\le q\le H, q\neq a,a'}\binom{n_{[q]}}{r(q)}, & \text{if } a \neq a', \\ \binom{n_{[a]}-2}{r(a)-1}\prod_{1\le q\le H, q\neq a}\binom{n_{[q]}}{r(q)}, & \text{if } a = a'. \end{cases}$$

Note that the total number of possible peers of units i and j is $\binom{n-2}{K-1}$. Because any possible peers of units i and j have the same probability, by symmetry,

$$\begin{aligned} \psi_{[a][a']}(r, r') &= \text{pr}(R_i = r, R_j = r' \mid j \in Z_i) \\ &= \begin{cases} \binom{n-2}{K-1}^{-1}\binom{n_{[a]}-1}{r(a)}\binom{n_{[a']}-1}{r(a')-1}\prod_{1\le q\le H, q\neq a,a'}\binom{n_{[q]}}{r(q)}, & \text{if } a' \in r, a \in r', r \setminus \{a'\} = r' \setminus \{a\}, a \neq a', \\ \binom{n-2}{K-1}^{-1}\binom{n_{[a]}-2}{r(a)-1}\prod_{1\le q\le H, q\neq a}\binom{n_{[q]}}{r(q)}, & \text{if } a' \in r, a \in r', r \setminus \{a'\} = r' \setminus \{a\}, a = a', \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

When units i and j are not in the same group, the total number of their possible peers is $\binom{n-2}{K}\binom{n-2-K}{K}$, and the total number of their possible peers such that units i and j receive treatments r and r' is

$$\begin{cases} \binom{n_{[a]}-1}{r(a)}\binom{n_{[a]}-1-r(a)}{r'(a)}\binom{n_{[a']}-1}{r(a')}\binom{n_{[a']}-1-r(a')}{r'(a')}\prod_{1\le q\le H, q\neq a,a'}\left\{\binom{n_{[q]}}{r(q)}\binom{n_{[q]}-r(q)}{r'(q)}\right\}, & \text{if } a \neq a', \\ \binom{n_{[a]}-2}{r(a)}\binom{n_{[a]}-2-r(a)}{r'(a)}\prod_{1\le q\le H, q\neq a}\left\{\binom{n_{[q]}}{r(q)}\binom{n_{[q]}-r(q)}{r'(q)}\right\}, & \text{if } a = a'. \end{cases}$$

Because any possible peers of units i and j have the same probability, by symmetry,

$$\begin{aligned} \phi_{[a][a']}(r, r') &= \text{pr}(R_i = r, R_j = r' \mid j \notin Z_i) \\ &= \begin{cases} \dfrac{\binom{n_{[a]}-1}{r(a)}\binom{n_{[a]}-1-r(a)}{r'(a)}\binom{n_{[a']}-1}{r(a')}\binom{n_{[a']}-1-r(a')}{r'(a')}\prod_{1\le q\le H, q\neq a,a'}\left\{\binom{n_{[q]}}{r(q)}\binom{n_{[q]}-r(q)}{r'(q)}\right\}}{\binom{n-2}{K}\binom{n-2-K}{3}}, & \text{if } a \neq a', \\ \dfrac{\binom{n_{[a]}-2}{r(a)}\binom{n_{[a]}-2-r(a)}{r'(a)}\prod_{1\le q\le H, q\neq a}\left\{\binom{n_{[q]}}{r(q)}\binom{n_{[q]}-r(q)}{r'(q)}\right\}}{\binom{n-2}{K}\binom{n-2-K}{K}}, & \text{if } a = a'. \end{cases} \end{aligned}$$

We have computed the four terms in (A14), and Theorem A1 follows directly. □