

Sensitivity analysis for contamination in egocentric-network randomized trials with interference

Bar Weinstein and Daniel Nevo*

Department of Statistics and Operations Research, Tel Aviv University

June 9, 2026

Abstract

Egocentric-Network Randomized Trials (ENRTs) are increasingly used to estimate causal effects under interference when measuring complete sociocentric network data is infeasible. ENRTs rely on egocentric network sampling, where a set of egos is first sampled, and each ego recruits a subset of its neighbors as alters. Treatments are then randomized across egos. While the observed ego-networks are disjoint by design, the underlying population network may contain edges connecting them, leading to contamination. Under a design-based framework, we show that the Horvitz-Thompson estimators of direct and indirect effects are biased whenever contamination is present. To address this, we derive bias-corrected estimators and propose a novel sensitivity analysis framework based on sensitivity parameters representing the probability or expected number of missing edges. This framework is implemented via both grid sensitivity analysis and probabilistic bias analysis, providing researchers with a flexible tool to assess the robustness of the causal estimators to contamination. We apply our methodology to the HIV Prevention Trials Network 037 study, finding that ignoring contamination may lead to underestimation of indirect effects and overestimation of direct effects.

Keywords: Causal inference; Network experiments; Network sampling; Probabilistic bias analysis; Spillovers.

1 Introduction

In causal inference and design of experiments, it is commonly assumed that there is no interference between units, that is, the treatment assigned to one unit does not affect the outcomes of other units (Cox, 1958). However, researchers are often interested in settings where the treatment of interest can spill over from treated units to untreated ones, leading to interference between units. For example, in infectious disease epidemiology, vaccinating some individuals may protect unvaccinated individuals (Halloran and Struchiner, 1991).

To estimate causal effects under interference, researchers often rely on detailed information about the network structure that encodes the relationships between units. Such data is typically assumed to be on a complete *sociocentric network*, where all units in the network and their connections are observed (Hudgens and Halloran, 2008; Aronow and Samii, 2017; Forastiere et al., 2021; Tchetgen Tchetgen et al., 2020; Ogburn et al., 2024). This

*barwein@tauex.tau.ac.il, danielnevo@tauex.tau.ac.il.

finite population can be composed of a collection of disjoint clusters or a single connected network. However, collecting complete sociocentric network data is often logistically challenging and expensive, limiting the application of network-based causal inference methods or resorting to sub-networks sampled from the population network.

A common sampling design is *egocentric network sampling* (Marsden, 2002; Perry et al., 2018), where a set of egos (also known as indexes) are sampled from the population, and each ego recruits a subset of its neighbors in the population network as alters (non-indexes). The observed network is then assumed to be composed of disjoint ego-networks comprising each ego and its recruited alters. This design is attractive due to its logistical feasibility and accessibility. Such egocentric sampling can be combined with randomized experiments, where treatments are randomly assigned to the recruited egos, resulting in an *egocentric-network randomized trial* (ENRT) (Buchanan et al., 2018, 2024; Chao et al., 2025; Fang et al., 2025). ENRTs are increasingly used in various studies, such as HIV prevention (Latkin et al., 2009, 2013; Booth et al., 2016; Miller et al., 2018), substance addiction (Sherman et al., 2009), and mental health (Asmus et al., 2017; Desrosiers et al., 2020).

However, while most methods assume that the ego-networks are disjoint, alters may be connected to multiple egos in the population network, and egos may be connected to other egos, resulting in contamination between ego-networks. Thus, the egocentric sampling process may lead to missing edges in the observed network, resulting in misspecification of the network interference structure (Weinstein and Nevo, 2026). Previous analyses of ENRTs typically assumed that the ego-networks are *non-overlapping* in the population network (Buchanan et al., 2018, 2024; Chao et al., 2025; Fang et al., 2025). That is, alters are connected only to one ego, and egos are not connected to other egos. Such an assumption may not hold in practice and be difficult to verify with the observed data alone, absent additional validation studies or reference data.

In this paper, we relax the non-overlapping networks assumption and allow for the possibility of overlapping ego-networks. Under a design-based inference paradigm, where the only source of randomness is the treatment assignment, we show that the Horvitz-Thompson (HT) estimators of the indirect and direct effects based on the observed ENRT data are biased whenever there is contamination between ego-networks. We develop a sensitivity analysis framework to assess the impact of such contamination on the causal estimates, and we couple it with bias-corrected estimators of the indirect and direct effects.

These bias-corrected estimators rely on sensitivity parameters that encode the probability or expected number of missing alter-ego and ego-ego edges in the population network. Crucially, because we treat the underlying population network as fixed but partially unobserved, these edge probabilities do not imply a stochastic superpopulation model of network generation. Rather, they quantify the researcher’s epistemic uncertainty regarding latent parts of the interference network. For the direct effect on egos, the framework requires an additional sensitivity parameter capturing the ratio of the sample average direct effect among egos exposed to at least one treated neighbor to that of unexposed egos. We provide practical guidelines for specifying these parameters under various scenarios using covariates and prior knowledge. Furthermore, while our framework is designed to evaluate robustness across various contamination scenarios, we also demonstrate how these sensitivity parameters can be directly calibrated using internal validation data, such as recall surveys in follow-up. This approach reduces the reliance on pure sensitivity analysis and can guide data collection strategies in future trials.

To improve asymptotic precision without compromising design-based validity, we augment these bias-corrected estimators with working outcome regression models. We further develop a conditional two-fold cross-fitting procedure tailored to the ENRT design. The

overall framework is implemented using either grid sensitivity analysis (Rosenbaum and Rubin, 1983; Greenland, 1996) or a probabilistic bias analysis (Greenland, 2005; Lash et al., 2014; Fox et al., 2021).

The HPTN 037 study (Latkin et al., 2009) serves as a primary motivating example for our work. HPTN 037 was an ENRT that aimed to evaluate the efficacy of a peer education intervention to reduce HIV risk behaviors among people who inject drugs. In the trial, treated egos received training on the prevention of risky behaviors and were encouraged to share this information with their alters. While previous analyses of this study (Latkin et al., 2009; Buchanan et al., 2018; Aroke et al., 2023; Chao et al., 2025) assumed that ego-networks are disjoint, contamination remains a significant concern. Specifically, alters might inject drugs or have sexual relations with egos from other ego-networks, or egos might be connected to each other. In this paper, we relax the assumption of disjoint ego-networks and assess the robustness of the causal estimates to contamination between the ego-networks. Our analysis suggests that ignoring contamination may lead to underestimation of the indirect effect and overestimation of the direct effect.

Our work relates to several strands of literature. First, it contributes to the growing literature on causal inference under interference where the true network structure is latent and only proxies are observed by the researchers (Toulis and Kao, 2013; Li et al., 2021; Hardy et al., 2019; Boucher and Houndetoungan, 2022; Weinstein and Nevo, 2025). Second, it adds to the literature on interference with sampled network data which has primarily focused on random node sampling or snowball sampling designs under linear-in-means parametric models (Chandrasekhar and Lewis, 2011; Sewell, 2017; Yauck, 2022; Marray, 2024). Third, it contributes to the literature on sensitivity analysis under interference, which has focused on unmeasured confounding (VanderWeele et al., 2014), or on settings where interference depends on two networks, but only one of which is observed (Egami, 2020). Finally, it relates to the literature on cross-cluster contamination in cluster-randomized trials, which often recommend changing the clusters boundaries to reduce contamination (Hayes and Moulton, 2017), or more recently, develop data-driven methods for clusters construction (Leung, 2025).

The rest of the paper is organized as follows. In Section 2, we introduce the setup of ENRTs, define the causal estimands of interest, and show that the commonly used HT estimators based on the observed data are biased under contamination. In Section 3, we develop a sensitivity analysis framework for contamination between ego-networks, providing bias-corrected estimators of the indirect and direct effects. In Section 4, we illustrate the proposed framework via simulations. In Section 5, we re-analyze the HPTN 037 study, assessing the robustness of the causal estimates to possible contamination. Finally, we conclude with a discussion in Section 6. All proofs are provided in the Appendix. The **R** package **ENRTsensitivity** that implements the proposed sensitivity analysis method is available at <https://github.com/barwein/ENRTsensitivity>.

2 Egocentric-network randomized trials

2.1 Setup and recruitment process

Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ denote the population network, where $\mathbf{V} = \{1, \dots, N\}$ is the set of units, and $\mathbf{E}$ is the set of edges, assumed to be binary. The size of the population N is assumed to be finite but unknown. The population network $\mathbf{G}$ can be described by its adjacency matrix $\mathbf{A}$, where $A_{ij} = 1$ if edge $(i, j) \in \mathbf{E}$, and 0 otherwise. We assume that the population network is undirected and that there are no self-edges, that is, $A_{ij} = A_{ji}$ and $A_{ii} = 0$ for

all i .

Units are recruited to the study via an *egocentric sampling design*. The sampling process consists of two stages. In the first stage, a set of egos (indexes) are sampled from the population. Then, in the second stage, each ego recruits a subset of its neighbors in the population network as alters (non-indexes). Formally, the recruitment process is as follows:

1. Sample egos from the population. Denote the set of recruited egos by $\mathcal{R}_e \subset \mathbf{V}$, and $n_e = |\mathcal{R}_e|$ as the number of egos. We do not impose any restrictions on the sampling process of egos.
2. Each ego recruits alters from its neighbors in the population network that were not recruited as egos in the first stage. We assume that each alter is recruited by a single ego, resulting in n_e disjoint ego-networks. This is often enforced in the recruitment process by researchers to avoid overlaps between the observed ego-networks, e.g., as was the case in the HPTN 037 study (Latkin et al., 2009). Denote the set of recruited alters by $\mathcal{R}_a \subset \mathbf{V} \setminus \mathcal{R}_e$, and let $n_a = |\mathcal{R}_a|$ be their number.

The set of all recruited units is $\mathcal{R} = \mathcal{R}_e \cup \mathcal{R}_a$, yielding an egocentric sample of size $n = n_e + n_a$. We assume that this sample is a strict subset of the population, i.e., $n < N$. The observed network $\tilde{\mathbf{G}} = (\mathcal{R}, \tilde{\mathbf{E}})$ is a subgraph of the population network $\mathbf{G}$ composed of n_e disjoint ego-networks. The set of observed edges is $\tilde{\mathbf{E}} = \{(i, j) \in \mathbf{E} : i \in \mathcal{R}_e, j \in \mathcal{R}_a, e(j) = i\}$, where $e(j) \in \mathcal{R}_e$ is the ego that recruited alter $j \in \mathcal{R}_a$. Clearly, $\tilde{\mathbf{E}} \subset \mathbf{E}$. Let $\tilde{\mathbf{A}}$ be the adjacency matrix of the observed network $\tilde{\mathbf{G}}$.

Figure 1 illustrates the egocentric sampling process in a small hypothetical population. Although the sampled egos are connected and some alters are connected to multiple egos in the population network, each ego reports only a subset of its true neighborhood, yielding an observed network made up of disjoint ego-networks.

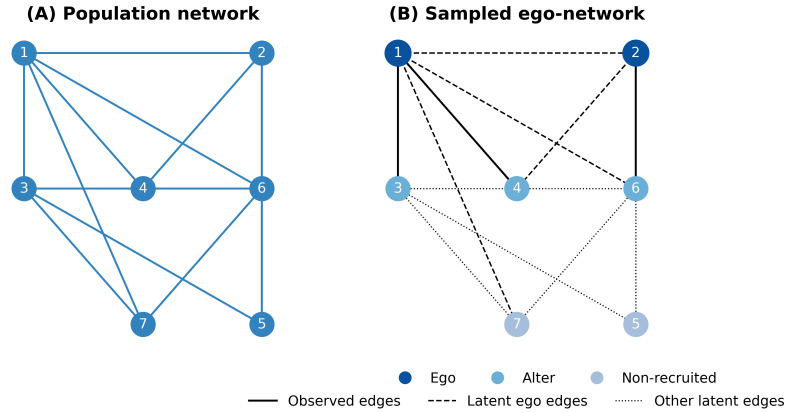


Figure 1: Illustration of egocentric sampling. Panel (A) shows the full $N = 7$ population network $\mathbf{G}$. Panel (B) illustrates a single egocentric sample $\tilde{\mathbf{G}}$: each ego (nodes 1 and 2; $n_e = 2$) reports only a subset of its true neighborhood as alters (nodes 3, 4, and 6; $n_a = 3$). The observed network is composed of disjoint ego-networks (solid lines), while ego-ego and unreported ego-alter connections remain latent (dashed lines). All other edges are also unobserved (dotted lines).

2.2 Experimental design, potential outcomes, and causal estimands

In what follows, all assumptions, estimands, and estimators are conditioned on the recruited sample. Specifically, we consider in-sample causal effects only. Let $\mathbf{Z}$ be the binary treat-

ment assignment vector, where $Z_i = 1$ if unit i is treated, and $Z_i = 0$ otherwise. We assume that treatments are randomly assigned to the recruited egos via Bernoulli randomization.

Assumption 1 (Bernoulli experimental design). *Treatments are independently assigned only to recruited egos, $\Pr(Z_i = 1 \mid \mathcal{R}) = p_z \mathbb{I}\{i \in \mathcal{R}_e\}$ with a known probability $p_z \in (0, 1)$.*

Denote the treatment space given the sample as $\mathcal{Z}_{\mathcal{R}} = \{\mathbf{z} \in \{0, 1\}^N : \Pr(\mathbf{Z} = \mathbf{z} \mid \mathcal{R}) > 0\}$, where treatments of non-recruited units are assumed to be equal to zero with probability one. To avoid clutter, the conditioning on $\mathcal{R}$ includes the specification of egos and alters sets as well. Furthermore, the distribution of $\mathbf{Z}$ always refers to the conditional distribution of $\mathbf{Z}$ given the recruitment $\mathcal{R}$.

Let $Y_i(\mathbf{z})$ be the potential outcomes of each unit under treatment vector $\mathbf{z} \in \mathcal{Z}_{\mathcal{R}}$. We take a design-based approach, where the potential outcomes are assumed to be fixed, and the only source of randomness is the treatment assignment. Let $\mathbf{Z}_{-i}$ denote the treatment assignment of all units except unit i . We assume that $\mathbf{Z}_{-i}$ affects the potential outcomes of unit i only through the values of the following exposure mapping (Manski, 2013; Aronow and Samii, 2017).

Assumption 2 (Exposure mapping). *Let $F(\mathbf{Z}_{-i}, \mathbf{A}) = \mathbb{I}\left\{\sum_{j \neq i} Z_j A_{ij} > 0\right\}$. For any $\mathbf{z}, \mathbf{z}' \in \mathcal{Z}_{\mathcal{R}}$, if $z_i = z'_i$ and $F(\mathbf{z}_{-i}, \mathbf{A}) = F(\mathbf{z}'_{-i}, \mathbf{A})$, then $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$.*

Denote the exposures in the true population network by $F_i = F(\mathbf{Z}_{-i}, \mathbf{A})$. The exposure F_i indicates whether at least one of the neighbors of unit i is treated. Assumption 2 implies that the exposure mapping is correctly specified (Sävje, 2024), and that only treatments of neighbors affect the potential outcomes of unit i . Binary exposure mapping specifications are common in the ENRT literature (Buchanan et al., 2018; Chao et al., 2025; Fang et al., 2025). For clarity, we present the main results under this binary specification, but extend our framework to a three-level exposure mapping specification distinguishing between zero, one, or two-or-more treated neighbors in Appendix B. By Assumption 2, we can express the potential outcomes as $Y_i(\mathbf{Z}) = Y_i(Z_i, F_i)$. Consequently, each unit has four effective potential outcomes, $Y_i(z, f)$ for $z, f \in \{0, 1\}^2$, indicating whether unit i is treated and whether at least one of its neighbors is treated. We assume that the observed outcomes Y_i , $i \in \mathcal{R}$ are generated from one of the potential outcomes.

Assumption 3 (Consistency). *For all $i \in \mathcal{R}$, the observed outcome is given by*

$$Y_i = \sum_{z, f \in \{0, 1\}^2} Y_i(z, f) \mathbb{I}\{Z_i = z, F_i = f\}.$$

In addition, we will also consider cases where covariates $\mathbf{X}_i$ are observed for each recruited unit. We will clarify their role later on.

We consider two types of causal estimands, separate for egos and alters. The first is the sample average indirect effect of treatment on the alters,

$$IE = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} [Y_i(0, 1) - Y_i(0, 0)]. \quad (1)$$

The IE is the average effect on alters of being exposed to at least one treated neighbor compared to not being exposed. In Appendix A.9, we also consider the indirect effect on the relative risk scale, for binary outcomes. The second estimand is the sample average direct effect of treatment on the egos,

$$DE = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} [Y_i(1, 0) - Y_i(0, 0)]. \quad (2)$$

The DE is the average effect on egos of being treated compared to not being treated when not exposed to any treated neighbors. Our explicit separation of causal estimands for egos and alters resolves interpretation issues of other estimands (Crawford et al., 2019). Furthermore, by not conflating the two groups, we avoid imposing additional assumptions and structure, as the recruitment process described in Section 2.1 may induce different selection mechanisms for egos and alters, leading to distinct potential outcome means across the two groups.

2.3 Naive estimation and bias under contamination

The observed data consist of the treatment assignments $\mathbf{Z}$, the outcomes Y_i , the observed network $\tilde{\mathbf{A}}$, and possibly covariates $\mathbf{X}_i$ of the recruited units $i \in \mathcal{R}$. By Assumption 3, the observed outcomes Y_i are generated from the potential outcomes with exposures F_i created by the *population network* $\mathbf{A}$. However, the observed network $\tilde{\mathbf{A}}$ might not include all the relevant edges present in the population network $\mathbf{A}$ due to the egocentric sampling process, resulting in a *misspecified network interference structure* (Weinstein and Nevo, 2026). Let $\tilde{F}_i = F(\mathbf{Z}_{-i}, \tilde{\mathbf{A}})$, be the observed exposures. Due to this network misspecification, the observed exposures $\tilde{F}_i$ may differ from the true exposures F_i . Specifically, under Assumption 2, only missing ego–ego and alter–ego edges induced by the sampling process can lead to incorrect observed exposures.

Consider, for example, the context of Figure 1. If only ego 1 is treated, for alter 6 we observe exposure $\tilde{F}_6 = 0$ since its ego in the observed network (node 2) is not treated. However, its true exposure is $F_6 = 1$ due to its link to ego 1 in the population network. Similarly, under the same treatment assignment, we will observe for ego 2 exposure $\tilde{F}_2 = 0$ while $F_2 = 1$ in the population network since ego 2 is connected to ego 1.

Due to the ENRT design, the observed exposures of egos are $\tilde{F}_i = 0$ for all $i \in \mathcal{R}_e$, while the observed exposures of alters can be either zero or one, depending on their observed ego’s treatment status. Specifically, by Assumption 1, alters are exposed to a treated ego in the observed network with probability $\Pr(\tilde{F}_i = 1 \mid i \in \mathcal{R}_a) = p_z$.

In practice, researchers estimate causal effects only with the observed exposures. The most common estimator in design-based inference settings is the Horvitz-Thompson (HT) estimator. For the indirect effect (1), its form is

$$\widehat{IE} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right], \quad (3)$$

and for the direct effect (2),

$$\widehat{DE} = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \left[\frac{\mathbb{I}\{Z_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{Z_i = 0\} Y_i}{1 - p_z} \right]. \quad (4)$$

If some alter–ego or ego–ego edges are missing in the observed network $\tilde{\mathbf{A}}$, both estimators (3) and (4) might be biased.

Proposition 1. *Under ENRT design and Assumptions 1–3,*

$$\begin{aligned} \mathbb{E}_{\mathbf{Z}} [\widehat{IE}] &= IE + \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \frac{p_z - \pi_i^a}{1 - p_z} [Y_i(0, 1) - Y_i(0, 0)], \\ \mathbb{E}_{\mathbf{Z}} [\widehat{DE}] &= DE + \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \pi_i^e \{ [Y_i(1, 1) - Y_i(0, 1)] - [Y_i(1, 0) - Y_i(0, 0)] \}, \end{aligned}$$

where $\pi_i^a = \Pr(F_i = 1 \mid i \in \mathcal{R}_a)$ and $\pi_i^e = \Pr(F_i = 1 \mid i \in \mathcal{R}_e)$ are the probabilities that an alter and an ego, respectively, are exposed to at least one treated neighbor in the population network $\mathbf{A}$.

When all alters are connected only to their own ego, then $\pi_i^a = p_z$ for all $i \in \mathcal{R}_a$, and $\widehat{IE}$ is an unbiased estimator of IE . Otherwise, whenever some alters are connected to more than one ego in the population network, $\pi_i^a > p_z$, and the indirect effect estimator (3) based on the observed exposures can be biased. The magnitude of the bias depends on the number of egos each alter is connected to in the population network. That is, the more egos an alter is connected to, the larger the gap between π_i^a and p_z , and hence the larger the bias. In addition, we show in Appendix A.5 that under unit-level monotonicity (in either direction) of the potential outcomes $Y_i(0, 1)$ and $Y_i(0, 0)$, the naive indirect estimator $\widehat{IE}$ is biased towards the null of no indirect effect.

If an ego is not connected to any other egos in the population network, then $\pi_i^e = 0$. If this is the case for all $i \in \mathcal{R}_e$, then $\widehat{DE}$ is an unbiased estimator of (2). However, if at least one ego i is connected to at least one other ego in the population network, then $\pi_i^e > 0$, and the direct effect estimator $\widehat{DE}$ based on the observed data can be biased. Similarly to alters, the magnitude of the bias depends on the number of ego-ego edges in the population network. Furthermore, in Appendix A.6, we analyze the direction of the bias of $\widehat{DE}$ relative to the null of no direct effect. We demonstrate that the direction of the bias is governed by the ratio of the sample average direct effect among egos exposed to at least one treated neighbor to the effect among non-exposed egos. Generally, if the direct effects are larger, on average, in the presence of treated neighbors than in their absence, $\widehat{DE}$ is biased away from the null. Conversely, if the direct effects are smaller, on average, in the presence of treated neighbors, the estimator is biased towards the null.

3 Sensitivity analysis for contamination

Proposition 1 shows that the HT estimators based on the observed data are biased whenever there is contamination between the ego-networks. We refer to contamination as the presence of alter-ego and ego-ego edges in the population network $\mathbf{A}$, that are missing from the observed network $\tilde{\mathbf{A}}$ due to the egocentric sampling process. Under such contamination, the ego-networks are not disjoint in the population network, leading to incorrect exposures when computed using the observed network.

By the structure of the egocentric sampling and the exposure mapping specification (Assumption 2), alters with observed exposure $\tilde{F}_i = 1$ will have correctly classified exposures. That is, $\tilde{F}_i = 1$ implies $F_i = 1$ with probability one for $i \in \mathcal{R}_a$. However, alters with $\tilde{F}_i = 0$ can have a true exposure of $F_i = 1$ or $F_i = 0$, depending on the structure of the population network $\mathbf{A}$. Therefore, the observed outcomes Y_i for alters with $\tilde{F}_i = 0$ can be generated from either $Y_i(0, 1)$ or $Y_i(0, 0)$.

On the other hand, all egos have observed exposures of $\tilde{F}_i = 0$, as they are linked only to alters in their ego-network in the observed network $\tilde{\mathbf{A}}$. However, the observed outcomes of egos that are assigned to treatment $Z_i = z$ can be generated from $Y_i(z, 1)$ or $Y_i(z, 0)$, implying that observed outcomes of egos are possibly linked to four different potential outcomes, compared to only two among the alters.

We develop a sensitivity analysis framework that enables researchers to assess the impact of possible contamination between ego-networks on the causal estimates. In Sections 3.1 and 3.2, we provide bias-corrected estimators of the indirect and direct effects that account for possible exposure misspecification due to the egocentric-network recruitment process.

Furthermore, in Section 3.3, we show how to augment these estimators with an outcome model through a cross-fitting procedure tailored to the design-based framework taken in this paper, enabling researchers to integrate covariates into the analysis to improve efficiency. Such model-assisted estimation is common in survey sampling (Särndal et al., 2003) and causal inference (Ding, 2024).

Building on Proposition 1, the bias-corrected indirect and direct effect estimators are functions of the exposure probabilities under the true population network π_i^a and π_i^e , respectively. The structure of the egocentric sampling design and our setup enables us to decompose π_i^a and π_i^e into basic components that are straightforward to compute given the specification of the presumed *expected number* or the *probability* of missing alter-ego and ego-ego edges. These components will form the basis of our sensitivity parameters, enabling seamless integration of covariates and prior knowledge into the sensitivity analysis. Crucially, we maintain the design-based framework throughout, treating the underlying population network $\mathbf{A}$ as fixed but partially unobserved. Therefore, probability statements regarding latent edges do not reflect a stochastic superpopulation model or assumptions about a data-generating process. Rather, they quantify the researcher’s epistemic uncertainty regarding the fixed, unobserved edges in the interference network. We provide multiple examples and practical guidelines on how to specify these parameters in Section 3.4.

For the direct effect, we require an additional sensitivity parameter to account for the complexity resulting from possibly four potential outcomes generating the observed outcomes. This parameter is the ratio of the sample average direct effect among egos exposed to at least one treated neighbor to the effect among non-exposed egos, as described in Section 3.2.

While the primary focus of this paper is to assess the robustness of the estimates over plausible contamination scenarios, researchers can sometimes calibrate the sensitivity parameters from auxiliary information. In Section 3.6, we demonstrate how different types of internal validation data, such as recall surveys at follow-up, can be leveraged to estimate these parameters. This approach reduces the reliance on pure sensitivity analysis and can motivate data collection and design choices in future studies.

3.1 Indirect effect

Define the bias-corrected indirect effect estimator as a function of the observed data and π_i^a ,

$$\widehat{IE}_{adj} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \frac{1 - p_z}{1 - \pi_i^a} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right]. \quad (5)$$

Given π_i^a , (5) is an unbiased estimator of the indirect effect (1).

Proposition 2. *Under ENRT design and Assumptions 1–3, we have $\mathbb{E}_{\mathbf{Z}}[\widehat{IE}_{adj}] = IE$.*

By the ENRT design and Assumption 2, $\pi_i^a \geq p_z$, with equality if and only if there are no additional latent alter-ego edges for alter i . Consequently, if all alters have the same exposure probability, $\pi_i^a = \pi^a$, then $\widehat{IE}_{adj} = \frac{1-p_z}{1-\pi^a} \widehat{IE}$. Therefore, $\text{sign}(\widehat{IE}_{adj}) = \text{sign}(\widehat{IE})$, and $\widehat{IE}_{adj} \geq \widehat{IE}$ if $\widehat{IE} > 0$ and vice versa otherwise.

We derive variance estimators for the bias-corrected estimator $\widehat{IE}_{adj}$. By the ENRT design, alters in the same ego-network all have the same treatment assignments and observed exposure status. Thus, we can calculate the variance by using the aggregated outcomes in each ego-network. Let $r_i^a = \frac{1-p_z}{1-\pi_i^a} \left[\frac{\mathbb{I}\{\tilde{F}_i=1\} Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i=0\} Y_i}{1-p_z} \right]$ such that $\widehat{IE}_{adj} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} r_i^a$.

The design-based variance estimator is (Ding, 2024)

$$\widehat{V}_{\mathbf{Z}} \left(\widehat{IE}_{adj} \right) = \frac{1}{n_a^2} \sum_{i \in \mathcal{R}_e} (T_i - \bar{T})^2, \quad (6)$$

where $T_i = \sum_{j \in \mathcal{R}_{a:e(j)=i}} r_j^a$ and $\bar{T} = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} T_i$ is the average over ego-networks. In Appendix A.7, we derive the variance estimator and establish the asymptotic normality of the bias-corrected estimator $\widehat{IE}_{adj}$, ensuring that Wald-type confidence intervals using the variance estimator $\widehat{V}_{\mathbf{Z}} \left(\widehat{IE}_{adj} \right)$ achieve a valid, albeit conservative, asymptotic coverage rate.

3.2 Direct effect

As the observed outcomes of egos are possibly generated from four different potential outcomes, compared to only two for the alters, bias correction for the direct effect estimator requires specifying the relation between the direct effect among non-exposed egos ($F_i = 0$) and the effect among egos exposed to at least one treated neighbor ($F_i = 1$). Denote the direct effect of treatment on an ego $i \in \mathcal{R}_e$ with exposure value $f \in \{0, 1\}$ by $\Delta_i(f) = Y_i(1, f) - Y_i(0, f)$. Let $\bar{\Delta}(f) = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \Delta_i(f)$ denote the sample average effect for $f \in \{0, 1\}$. That is, with a slight abuse of notations, $\bar{\Delta}(0) = DE$ as defined in (2). Let κ be the ratio of the sample average direct effect among egos exposed to at least one treated neighbor $\bar{\Delta}(1)$ to the effect among non-exposed egos $\bar{\Delta}(0)$:

$$\bar{\Delta}(1) = \kappa \bar{\Delta}(0). \quad (7)$$

Under the design-based framework taken in the paper, where the potential outcomes are fixed, there is always a constant κ satisfying (7). The parameter κ will serve as an additional sensitivity analysis parameter. We discuss the interpretation of κ below.

Let $\bar{\pi}^e = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \pi_i^e$ be the sample average probability of exposure among the egos. We define an inflation factor $u_e = n_e [1 + \bar{\pi}^e(\kappa - 1)]$ that adjusts the effective sample size of egos based on the expected contamination and κ . Using this factor, we define the bias-corrected direct effect estimator as a function of the observed data, $\bar{\pi}^e$ and κ ,

$$\widehat{DE}_{adj} = \frac{1}{1 + \bar{\pi}^e(\kappa - 1)} \widehat{DE} = \frac{1}{u_e} \sum_{i \in \mathcal{R}_e} \left[\frac{\mathbb{I}\{Z_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{Z_i = 0\} Y_i}{1 - p_z} \right]. \quad (8)$$

An additional assumption is needed for $\widehat{DE}_{adj}$ to be unbiased, as we now formalize. For two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$, let the empirical covariance be $S_{\mathbf{a}, \mathbf{b}} = \frac{1}{m} \sum_{i=1}^m (a_i - \bar{a})(b_i - \bar{b})$, where $\bar{a} = \frac{1}{m} \sum_i a_i$. To eliminate bias, we require that the empirical covariance between exposure probabilities π_i^e and the direct effect among exposed egos $\Delta_i(1)$ is equal to the empirical covariance between π_i^e and the direct effect among non-exposed egos $\Delta_i(0)$. Let $\boldsymbol{\pi}^e = (\pi_1^e, \dots, \pi_{n_e}^e)$ and $\boldsymbol{\Delta}_i(f) = (\Delta_1(f), \dots, \Delta_{n_e}(f))$ be the vectors of exposure probabilities and direct effects at the unit-level of the egos.

Assumption 4. $S_{\boldsymbol{\pi}^e, \boldsymbol{\Delta}(1)} = S_{\boldsymbol{\pi}^e, \boldsymbol{\Delta}(0)}$.

To provide a better understanding of the implications of Assumption 4, consider the representation of the potential outcomes under a saturated model

$$Y_i(z, f) = \beta_{0i} + \beta_{1i}z + \beta_{2i}f + \beta_{3i}zf.$$

Under this representation, $\Delta_i(1) = \beta_{1i} + \beta_{3i}$, $\Delta_i(0) = \beta_{1i}$, and simple calculations show that Assumption 4 is equivalent to

$$S_{\pi^e, \beta_3} = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} (\pi_i^e - \bar{\pi}^e)(\beta_{3i} - \bar{\beta}_3) = 0. \quad (9)$$

That is, the empirical covariance between the exposure probabilities π_i^e and the treatment-exposure interaction terms β_{3i} is zero. Eq. (9) naturally holds if these interaction effects are homogeneous across all egos (i.e., $\beta_{3i} = \beta_3$ for some constant β_3) or if the exposure probabilities are homogeneous (i.e., $\pi_i^e = \pi^e$ for some constant $\pi^e \in [0, 1]$). In practice, this assumption would only be violated if egos that are highly connected to other egos systematically exhibit different interaction responses to the intervention compared to less connected egos.

Given π_i^e and κ , (8) is an unbiased estimator of the direct effect (2).

Proposition 3. *Under ENRT design and Assumptions 1–4, we have $\mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] = DE$.*

As described above, under homogeneous exposure probabilities, the empirical covariance requirement stated in Assumption 4 is trivially satisfied.

Corollary 1. *If exposure probabilities are homogeneous ($\pi_i^e = \pi^e$ for all $i \in \mathcal{R}_e$), then $\mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] = DE$.*

Reasoning about the range of plausible κ values depends on the mechanism under study. For example, in vaccine efficacy studies, if treatment represents vaccines, it is reasonable to assume that the direct effect of vaccination is larger, on average, among egos with non-vaccinated neighbors compared to those with vaccinated neighbors (who benefit from some indirect protection), implying $\kappa < 1$. On the other hand, in the HPTN 037 study, which we analyze in Section 5, the treatment is a behavioral intervention reliant on peer education. In such settings, researchers may believe the intervention to be mutually reinforcing. That is, the direct effect may be larger, on average, for egos who are also exposed to at least one treated ego compared to those who are not, which implies $\kappa > 1$. While researchers must often reason about ranges for κ based on subject-matter expertise, it is also possible to estimate κ directly if appropriate internal validation data are collected, as we detail in Section 3.6.

If $\pi_i^e = 0$ or $\kappa = 1$, the bias-corrected estimator $\widehat{DE}_{adj}$ reduces to the naive HT estimator $\widehat{DE}$. Thus, $\widehat{DE}$ is unbiased if, for example, $\kappa = 1$ and the probabilities π_i^e are homogeneous. That is, $\widehat{DE}$ can be an unbiased estimator of DE even if there is some contamination between the egos ($\pi^e > 0$). See Appendix A.4 for more details.

Furthermore, if $\kappa > 1 - \frac{1}{\bar{\pi}^e}$ (e.g., for $\bar{\pi}^e = 0.01$ this implies $\kappa > -99$), $\text{sign}(\widehat{DE}_{adj}) = \text{sign}(\widehat{DE})$, i.e., the bias-corrected estimator will have the same sign as the naive estimator. In addition, if $\kappa > 1$ or $\kappa \in (1 - \frac{1}{\bar{\pi}^e}, 1)$, the bias-corrected estimator $\widehat{DE}_{adj}$ is closer to or further away from, respectively, zero than the naive estimator $\widehat{DE}$. See Appendix A.6 for further details.

The requirement of Assumption 4 in Proposition 3 can be replaced with alternative formulations. We derive nonparametric bounds for DE based on the observed data, π_i^e and κ , enabling researchers to obtain valid intervals for DE without Assumption 4. Furthermore, we show that replacing the average-level proportionality between $\bar{\Delta}(1)$ and $\bar{\Delta}(0)$, as expressed in (7), with a unit-level proportionality assumption between $\Delta_i(1)$ and $\Delta_i(0)$,

produces a modified bias-corrected estimator that is an unbiased estimator of DE without Assumption 4. See Appendix A.10 for more details.

Our simulation study (Section 4 and Appendix D) assessed the impacts of violating Assumption 4 and misspecifying the sensitivity parameter κ . We found that violations of Assumption 4 resulted in negligible bias for the corrected direct effect estimators, albeit with slight variance underestimation. The impact of misspecifying κ , however, depends on its direction. If κ is misspecified in the wrong direction (e.g., specifying $\kappa < 1$ when the true $\kappa > 1$), the bias-corrected estimator can exhibit greater bias and lower coverage than the naive estimator. Conversely, when κ is misspecified in magnitude but correctly oriented in direction, the bias-corrected estimator outperforms the naive estimator, maintaining smaller bias and higher coverage.

We derive a conservative variance estimator for the bias-corrected estimator $\widehat{DE}_{adj}$. Let $r_i^e = \frac{\mathbb{I}\{Z_i=1\}Y_i}{p_z} - \frac{\mathbb{I}\{Z_i=0\}Y_i}{1-p_z}$ such that $\widehat{DE}_{adj} = \frac{1}{u_e} \sum_{i \in \mathcal{R}_e} r_i^e$. By the ENRT design, treatment assignment is independent across egos. The Neyman-type variance estimator is therefore (Ding, 2024),

$$\widehat{V}_{Neyman}(\widehat{DE}_{adj}) = \frac{1}{u_e^2} \sum_{i \in \mathcal{R}_e} (r_i^e - \widehat{DE}_{adj})^2.$$

This estimator assumes that the terms r_i^e are independent. However, in the presence of contamination, this independence assumption is violated. If two egos i and j share a neighbor $k \in \mathcal{R}_e$ in the population network $\mathbf{A}$, then the terms r_i^e and r_j^e are not independent. This dependence induces pairwise covariances ($\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e)$) that are omitted from $\widehat{V}_{Neyman}(\widehat{DE}_{adj})$. To ensure valid inference, we propose a conservative estimator $\widehat{V}_{Conta.}(\widehat{DE}_{adj})$ for the variance in $\widehat{DE}_{adj}$ attributed to contamination between ego-networks, derived using the edge probabilities under the sensitivity model (presented in Section 3.4) to account for these latent correlations. The final corrected variance estimator for $\widehat{DE}_{adj}$ is given by

$$\widehat{V}_{\mathbf{Z}}(\widehat{DE}_{adj}) = \widehat{V}_{Neyman}(\widehat{DE}_{adj}) + \widehat{V}_{Conta.}(\widehat{DE}_{adj}). \quad (10)$$

In Appendix A.7 we provide the derivation of the variance estimator $\widehat{V}_{\mathbf{Z}}(\widehat{DE}_{adj})$ and, analogously to Section 3.1, show the asymptotic normality of $\widehat{DE}_{adj}$ and that Wald-type confidence intervals using the variance estimator $\widehat{V}_{\mathbf{Z}}(\widehat{DE}_{adj})$ achieve a valid, albeit conservative, asymptotic coverage rate.

3.3 Outcome-model augmented estimators

The bias-corrected estimators $\widehat{IE}_{adj}$ and $\widehat{DE}_{adj}$ can be improved by incorporating outcome regression models that leverage information from the covariates $\mathbf{X}_i$. Crucially, the incorporation of these models does not constitute a departure from our design-based paradigm. The regression models serve strictly as working models used to improve asymptotic precision, and the theoretical validity of the resulting augmented estimators relies solely on the randomization of the treatment assignment (Lin, 2013; Abadie et al., 2020; Gao and Ding, 2025). As we demonstrate in our simulation study (Section 4), this augmentation yields valid variance estimates, overcoming the conservative variance estimation of the non-augmented estimators, while maintaining negligible bias.

To fully preserve this design-based validity, standard regression adjustments require careful implementation. While sample splitting (i.e., cross-fitting) is the standard solution

to avoid finite-sample bias, standard splitting schemes can violate the design-based independence assumptions (Lu et al., 2025). To address this, we employ a conditional two-fold cross-fitting procedure tailored to the ENRT design. Specifically, we adapt Algorithm 1 of Lu et al. (2025) to our setting. We describe here the algorithm for completeness, but refer readers to Lu et al. (2025) for additional details. We focus on the augmented bias-corrected estimators, but the naive estimators $\widehat{IE}$ and $\widehat{DE}$ can also be augmented similarly, albeit without the bias-correction weighting.

Let $\widehat{\mu}_i^a(f)$ denote the predicted value for alter i under exposure $\widetilde{F}_i = f$, from an outcome model estimated using covariates $\mathbf{X}_i$. For example, $\widehat{\mu}_i^a(f)$ can be the predicted value from a linear model of the form $Y_i \sim \widetilde{F}_i + \mathbf{X}_i$. Similarly, let $\widehat{\mu}_i^e(z)$ be the predicted value for ego i under treatment $Z_i = z$, from an outcome model estimated using covariates $\mathbf{X}_i$. The two-fold cross-fitting algorithm is as follows.

1. Randomly split the observed ego-networks into two disjoint parts S_0, S_1 such that $|S_0 \cup S_1| = n$ and $S_0 \cap S_1 = \emptyset$. Specifically, we assign each ego-network (ego and alters) into S_1 with probability 0.5.
2. For $q = 0, 1$:
 - (a) Estimate the outcome models for alters and egos using data from the complementary split S_{1-q} .
 - (b) Predict $\widehat{\mu}_i^a(f)$ and $\widehat{\mu}_i^e(z)$ on the units from S_q .
 - (c) Estimate IE on units from S_q :

$$\widehat{IE}_{adj[q]}^{aug} = \frac{1}{n_{a[q]}} \sum_{i \in \mathcal{R}_a \cap S_q} \frac{1 - p_z}{1 - \pi_i^a} [D_i^a + \widehat{\mu}_i^a(1) - \widehat{\mu}_i^a(0)],$$

where $n_{a[q]} = \sum_{i \in \mathcal{R}_a} \mathbb{I}\{i \in S_q\}$ is the number of alters in split S_q , and where

$$D_i^a = \frac{\mathbb{I}\{\widetilde{F}_i = 1\} (Y_i - \widehat{\mu}_i^a(1))}{p_z} - \frac{\mathbb{I}\{\widetilde{F}_i = 0\} (Y_i - \widehat{\mu}_i^a(0))}{1 - p_z}.$$

- (d) Estimate DE on units from S_q :

$$\widehat{DE}_{adj[q]}^{aug} = \frac{1}{u_{e[q]}} \sum_{i \in \mathcal{R}_e \cap S_q} [D_i^e + \widehat{\mu}_i^e(1) - \widehat{\mu}_i^e(0)],$$

where $u_{e[q]} = n_{e[q]}[1 + \bar{\pi}^{e[q]}(\kappa - 1)]$ similarly to u_e in Section 3.2, with $n_{e[q]} = \sum_{i \in \mathcal{R}_e} \mathbb{I}\{i \in S_q\}$ and $\bar{\pi}^{e[q]} = \frac{1}{n_{e[q]}} \sum_{i \in \mathcal{R}_e \cap S_q} \pi_i^e$ are the number of egos and the average exposure probability, respectively, in split S_q , and

$$D_i^e = \frac{\mathbb{I}\{Z_i = 1\} (Y_i - \widehat{\mu}_i^e(1))}{p_z} - \frac{\mathbb{I}\{Z_i = 0\} (Y_i - \widehat{\mu}_i^e(0))}{1 - p_z}.$$

3. Combine the estimates across splits

$$\begin{aligned} \widehat{IE}_{adj}^{aug} &= \sum_{q=0,1} \frac{n_{a[q]}}{n_a} \widehat{IE}_{adj[q]}^{aug}, \\ \widehat{DE}_{adj}^{aug} &= \sum_{q=0,1} \frac{u_{e[q]}}{u_e} \widehat{DE}_{adj[q]}^{aug}. \end{aligned} \tag{11}$$

Variance estimation is a combination across splits of variance estimators (6) for the IE estimator and (10) for the DE estimator (Lu et al., 2025). In Appendix A.8, we show that the augmented estimators $\widehat{IE}_{adj}^{aug}$ and $\widehat{DE}_{adj}^{aug}$ remain unbiased estimators of IE and DE , respectively, derive their variance estimator, and establish their asymptotic normality, ensuring that Wald-type confidence intervals based on their associated variance estimators achieve at least the nominal asymptotic coverage rate.

3.4 Specifying the sensitivity parameters

Propositions 2 and 3 establish that the bias-corrected estimators of the indirect and direct effects depend on the unknown exposure probabilities π_i^a and π_i^e of alters and egos, respectively, in the population network $\mathbf{A}$. For the direct effect, the bias-corrected estimator also depends on the sensitivity parameter κ . We now illustrate how the computation of π_i^a and π_i^e reduces to specification of edge-level probabilities or the expected number of missing alter–ego and ego–ego edges. These specifications form the basis of our sensitivity parameters. Importantly, while we maintain the design-based framework, implying that the population network $\mathbf{A}$ is fixed, its full structure is unobserved. Consequently, we treat the presence of missing edges as uncertain for the purpose of the sensitivity analysis. We do not posit a generative superpopulation model for $\mathbf{A}$. Rather, we define sensitivity parameters to model our uncertainty about the missing alter–ego and ego–ego edges. Therefore, all subsequent probability statements regarding missing edges should be interpreted strictly as quantifications of this uncertainty.

By the exposure mapping specification (Assumption 2) and the experimental design (Assumption 1), the relevant edges in the population network $\mathbf{A}$ that determine the probabilities π_i^a and π_i^e are only the alter–ego or ego–ego edges. In the observed network $\tilde{\mathbf{A}}$, each alter $i \in \mathcal{R}_a$ is connected only to its ego $e(i) \in \mathcal{R}_e$. Let $\mathcal{R}_e^j = \mathcal{R}_e \setminus \{j\}$ be the set of all egos other than $j \in \mathcal{R}_e$. The latent alter–ego edges $\{A_{ij} : i \in \mathcal{R}_a, j \in \mathcal{R}_e^{e(i)}\}$, and ego–ego edges $\{A_{ij} : i, j \in \mathcal{R}_e, i \neq j\}$ are the edges missing from $\tilde{\mathbf{A}}$, relevant to the exposure probabilities. Let ρ_{ij}^e and ρ_{ij}^a represent the postulated probabilities that a missing edge exists, i.e.,

$$\begin{aligned}\rho_{ij}^e &= \Pr(A_{ij} = 1 \mid i, j \in \mathcal{R}_e, i \neq j), \\ \rho_{ij}^a &= \Pr(A_{ij} = 1 \mid i \in \mathcal{R}_a, j \in \mathcal{R}_e^{e(i)}).\end{aligned}\tag{12}$$

Using (12), we can derive expressions for π_i^e and π_i^a . By the law of total probability and Assumption 1, $\Pr(Z_j A_{ij} = 1 \mid i \neq j \in \mathcal{R}_e) = p_z \rho_{ij}^e$. Assuming edge independence yields

$$\pi_i^e = 1 - \prod_{j \in \mathcal{R}_e^i} (1 - p_z \rho_{ij}^e).\tag{13}$$

Turning to π_i^a , alter $i \in \mathcal{R}_a$ is exposed ($F_i = 1$) only if its ego $e(i)$ or at least one of the other recruited egos connected to i in the population network $\mathbf{A}$ is treated. Therefore, we obtain (Appendix A.2)

$$\pi_i^a = p_z + (1 - p_z) \left[1 - \prod_{j \in \mathcal{R}_e^{e(i)}} (1 - p_z \rho_{ij}^a) \right].\tag{14}$$

Eqs. (13) and (14) illustrate how the exposure probabilities can be computed given specifications of the edge-level sensitivity parameters ρ_{ij}^e and ρ_{ij}^a .

We now discuss some prominent examples of how researchers can specify these parameters in practice. The examples include both homogeneous and heterogeneous edge probabilities, as well as specifications based on the expected number of missing edges.

Example 1 (Homogeneous probabilities). If researchers are willing to specify the same probabilities (12) for all units, i.e., $\rho_{ij}^e = \rho^e$ and $\rho_{ij}^a = \rho^a$, then the exposure probabilities (13) and (14) simplify to

$$\begin{aligned}\pi_i^e &= 1 - (1 - p_z \rho^e)^{n_e - 1}, \\ \pi_i^a &= p_z + (1 - p_z) [1 - (1 - p_z \rho^a)^{n_e - 1}].\end{aligned}$$

For instance, if researchers suspect that about 1% of the ego-ego edges are possible, they can set $\rho^e = 0.01$.

Example 2 (Homogeneous number of edges). A special case of Example 1 occurs by specifying the approximate *expected number* of latent edges, separately for the ego-ego and alter-ego edges. Let $m^e \in \{1, \dots, \binom{n_e}{2}\}$ be the total number of ego-ego edges that are believed to be missing. Setting

$$\rho_{ij}^e = \rho^e(m^e) = \frac{m^e}{\binom{n_e}{2}},$$

weights equally each ego-ego edge probability and yields m^e expected ego-ego edges. There are $n_a(n_e - 1)$ possible missing alter-ego edges. If approximately m^a such edges are expected to be missing, set the probability of each edge to

$$\rho_{ij}^a = \rho^a(m^a) = \frac{m^a}{n_a(n_e - 1)}.$$

For instance, if researchers suspect that each alter is connected to approximately $c^a \geq 0$ additional egos, they can select $m^a = c^a \times n_a$. Similarly, if each ego is believed to be connected to approximately $c^e \geq 0$ additional egos, then $m^e = \frac{c^e \times n_e}{2}$.

Example 3 (Heterogeneous probabilities). Example 1 can be extended by allowing edge probabilities to vary based on unit-level covariates $\mathbf{X}_i$. This captures the well-known phenomenon in network science of homophily (McPherson et al., 2001), where units with similar characteristics are more likely to interact.

We begin with a baseline probability parameter, ρ_{base}^e , analogous to the homogeneous parameter ρ^e in Example 1. We then adjust this baseline using a pairwise dissimilarity measure $d_{ij} = d(\mathbf{X}_i, \mathbf{X}_j)$. Common choices include the L^p norm, $d_{ij} = \|\mathbf{X}_i - \mathbf{X}_j\|_p$, or the Cosine distance, $d_{ij} = 1 - \frac{\mathbf{X}_i^T \mathbf{X}_j}{\|\mathbf{X}_i\| \|\mathbf{X}_j\|}$. To incorporate these distances while preserving symmetry ($\rho_{ij}^e = \rho_{ji}^e$), we represent the edge probabilities using the logistic link function:

$$\text{logit}(\rho_{ij}^e) = \text{logit}(\rho_{base}^e) - \gamma^e (d_{ij} - \bar{d}^e),$$

where $\bar{d}^e = \frac{2}{n_e(n_e - 1)} \sum_{i < j \in \mathcal{R}_e} d_{ij}$ is the average distance between all ego pairs. Here, $\gamma^e \geq 0$ is a sensitivity parameter controlling the *strength of homophily*. By centering the distances (subtracting $\bar{d}^e$), we decouple the network density from the homophily structure such that ρ_{base}^e controls the overall network density and γ^e controls how strongly edges concentrate

among similar units. Setting $\gamma^e = 0$ eliminates the distance term, recovering the homogeneous probabilities of Example 1 ($\rho_{ij}^e = \rho_{base}^e$). Larger values of γ^e decrease the probability of edges between dissimilar units (where $d_{ij} > \bar{d}^e$) and increase it for similar units.

Analogous definitions apply for alter-ego edges using a baseline ρ_{base}^a , a sensitivity parameter γ^a , and centered distances between alters and potential egos.

Example 4 (Heterogeneous number of missing edges). Examples 2 and 3 can be combined to allow pair-specific heterogeneous probabilities while fixing the expected number of missing edges. Let $w_{ij}^e = \exp(-\gamma^e d_{ij}^e)$ and $w_{ij}^a = \exp(-\gamma^a d_{ij}^a)$ be the exponential of the negative measures defined in Example 3, for fixed values of γ^e and γ^a . Larger values of w_{ij}^e and w_{ij}^a indicate higher similarity between the covariates of i and j . Define the total sum of weights among all distinct ego-ego and alter-ego pairs with missing edges:

$$W^e = \sum_{i,j \in \mathcal{R}_e, i < j} w_{ij}^e, \quad W^a = \sum_{i \in \mathcal{R}_a} \sum_{j \in \mathcal{R}_e^{(i)}} w_{ij}^a.$$

Similar to Example 2, researchers can specify m^e and m^a , the approximate number of missing ego-ego and alter-ego edges, respectively, and set:

$$\rho_{ij}^e = \frac{m^e}{W^e} w_{ij}^e, \quad \rho_{ij}^a = \frac{m^a}{W^a} w_{ij}^a.$$

This specification reflects that the expected number of missing edges is approximately m^e and m^a , while allowing for heterogeneous edge probabilities. Setting all weights (w_{ij}^e and w_{ij}^a) to one recovers Example 2.

3.5 Running the sensitivity analysis in practice

To assess the impact of contamination on the causal estimates, researchers can utilize the bias-corrected estimators for the indirect effect and for the direct effect. The bias-corrected estimators depend on the sensitivity parameters that define the edge-level probabilities (12) and the parameter κ for the direct effect. Examples 1-4 provide practical guidelines on how to specify these parameters in practice. When the true values of such parameters are unknown and cannot be estimated from the data, researchers can resort to sensitivity analysis by exploring a range of plausible values. We consider here two approaches: grid sensitivity analysis (GSA) and probabilistic bias analysis (PBA).

In GSA, researchers specify a grid of values for the sensitivity parameters and compute the bias-corrected estimates for each combination of parameters (Rosenbaum and Rubin, 1983; Greenland, 1996). That results in a grid of bias-corrected estimates. Such estimates can be coupled with the associated confidence interval given each value of the sensitivity parameters in the specified grid. Researchers can then summarize the sensitivity analysis results by reporting the answers for questions such as “what is the largest number of missing alter-ego edges such that the confidence interval for IE excludes the value zero?” or “are there combinations of ρ_{ij}^e and κ such that $\widehat{DE}_{adj}$ flips its sign compared to the naive estimator?”.

In PBA, researchers specify a distribution for the sensitivity parameters and compute the distribution of bias-corrected estimates over this distribution (Greenland, 2005; Lash et al., 2014; Fox et al., 2021). In practice, this distribution is approximated using Monte Carlo by sampling from the specified distribution, and summarized through averages and

intervals of the resulting bias-corrected estimates. This distribution of bias-corrected estimates accounts for uncertainty in the sensitivity parameters (bias uncertainty). Furthermore, as described below, it is possible to compute a distribution of estimates that accounts for both statistical uncertainty and bias uncertainty. In our design-based framework, statistical uncertainty arises solely from the treatment assignment distribution. A review of PBA is given in Appendix C.

We now describe the GSA and PBA approaches in more details.

Grid sensitivity analysis. Given a grid of sensitivity parameters, the full procedure is as follows:

1. (Optional.) Get the predicted values for the outcome models $\hat{\mu}_i^a(f)$ and $\hat{\mu}_i^e(z)$ using the cross-fitting procedure described in Section 3.3. These predicted values are calculated only once and used for all sensitivity parameter values.
2. For each parameter value in the grid of sensitivity parameters:
 - (a) Compute bias-corrected estimates $\widehat{IE}_{adj}$ or $\widehat{IE}_{adj}^{aug}$ and $\widehat{DE}_{adj}$ or $\widehat{DE}_{adj}^{aug}$.
 - (b) Estimate the variance of each bias-corrected estimator.
 - (c) Compute confidence intervals for each estimator.

The results are then summarized by presenting the estimators and associated confidence intervals as a function of the sensitivity parameter(s), as we demonstrate in the data analysis (Section 5). Details on estimating variance and confidence intervals are provided in Sections 3.1-3.3.

Probabilistic bias analysis. Instead of specifying a grid of sensitivity parameters, researchers assign a distribution for each parameter. For instance, under the setting in Example 1, if researchers believe that approximately 1% of the possible ego-ego edges are missing from $\tilde{\mathbf{A}}$, they can set $\rho^e \sim \text{Beta}(a, b)$ and choose the values of a and b such that the mean is $\frac{a}{a+b} = 0.01$, and the variance reflects the level of uncertainty about ρ^e . Let $p(\cdot)$ be the specified distribution for the sensitivity parameters. In practice, researchers can specify a separate distribution for each parameter.

Let $B > 0$ be the number of Monte Carlo samples. The PBA procedure that accounts for both bias and design uncertainty proceeds as follows.

1. For each $b = 1, \dots, B$:
 - (a) Sample sensitivity parameters from the distribution $p(\cdot)$.
 - (b) Compute the bias-corrected estimates $\widehat{IE}_{adj}^{(b)}$ and $\widehat{DE}_{adj}^{(b)}$ under the sampled values.
 - (c) Draw from the (approximate) distribution of the estimator

$$\widehat{IE}_{adj}^{(b,rand)} \sim N \left(\widehat{IE}_{adj}^{(b)}, \hat{V}_{\mathbf{Z}} \left(\widehat{IE}_{adj} \right) \right),$$

and similarly for the other estimators.

2. Summarize the distribution of bias-corrected estimates. For example, by computing the average $\frac{1}{B} \sum_{b=1}^B \widehat{IE}_{adj}^{(b,rand)}$ and empirical percentiles for intervals.

If researchers use the augmented estimators $\widehat{IE}_{adj}^{aug}$ and $\widehat{DE}_{adj}^{aug}$ (Section 3.3), the outcome models can be pre-computed as in GSA step 1. Step 1c propagates the statistical uncertainty (w.r.t. $\mathbf{Z}$) into the estimators by simulating a draw from the estimators' asymptotic normal distribution. Therefore, the distribution of estimates $\widehat{IE}_{adj}^{(b,rand)}$ accounts for both bias uncertainty about the sensitivity parameters and statistical uncertainty resulting from the treatments assignment (Greenland, 2005; Fox et al., 2021). Further technical details and motivation of the PBA procedure are provided in Appendix C.

3.6 Estimation of sensitivity parameters with internal validation data

Until this section, the exposure probabilities π_i^a and π_i^e (or the basic components of latent edge probabilities ρ_{ij}^a and ρ_{ij}^e) and κ have been treated as unknown sensitivity parameters. However, these parameters can be directly estimated with different types of internal validation data, thus alleviating the reliance on pure sensitivity analysis. We now describe the specific data types required for estimating these parameters, separately for alters (network members) and egos.

Internal validation data for alters. Recall that by ENRT design and the exposure mapping specification (Assumption 2), alters' exposure misclassification is one-sided. That is, alters with observed exposure $\tilde{F}_i = 1$ will have a correctly classified exposure, while those with $\tilde{F}_i = 0$ can have a true exposure of $F_i = 0$ or $F_i = 1$. Therefore, in the study follow-up period, researchers can use a validation subset of alters from the non-treated ego-networks and ask them to recall information about the intervention. For example, in the HPTN 037 study, the treated egos underwent training sessions that included a number of specific phrases and terms designed to assess the diffusion of intervention messages. Thus, alters connected to a non-treated ego should have no recollection of these terms. In the follow-up period, alters within a validation subset were asked to recall terms associated with the intervention to assess possible contamination (Simmons et al., 2015; Aroke et al., 2023; Chao et al., 2025). This information can be used to estimate the conditional probability $\Pr(F_i = 1 \mid \tilde{F}_i = 0, i \in \mathcal{R}_a)$ and subsequently the exposure probabilities π_i^a or, in the case of homogeneous probabilities (Examples 1-2), the edge probabilities ρ^a . We describe this in more detail in Appendix A.11. In Section 5, we compare the bias-adjusted indirect effect estimates resulting from the calibration using the internal validation data of the study to the sensitivity analysis described in the preceding subsections.

Internal validation data for egos. Similar recall data described for the alters can be used to assess the exposure probabilities of egos, in the case of homogeneous exposure probabilities. Non-treated egos should have no recollection of the terms described in the intervention. Therefore, non-treated egos that recall intervention-related information in the follow-up period ostensibly obtained this information due to contamination from other egos. Under homogeneous exposure probabilities, both π^e and ρ^e can be estimated using the proportion of non-treated egos in the follow-up period that recalled intervention-related information (Appendix A.11).

Alternatively, researchers can ask the egos more direct and informed questions. For example, instead of recalling *any* terms associated with the intervention, researchers can ask the egos if they learned about the intervention from other friends or sources. That will provide information regarding the true exposure status of egos in the validation subsample. These internal validation data can be used to estimate both π_i^e and κ (Appendix A.11).

This type of information was not collected in the HPTN 037 study and can be of valuable importance for future studies.

4 Simulation study

We conducted a simulation study to evaluate the performance of the bias-corrected estimators $\widehat{IE}_{adj}$, $\widehat{IE}_{adj}^{aug}$, $\widehat{DE}_{adj}$, and $\widehat{DE}_{adj}^{aug}$. We assessed the bias, variance estimation, and empirical coverage of the confidence intervals under the true values of the sensitivity parameters. Furthermore, we evaluated how misspecification of κ values or violation of Assumption 4 affects the estimation of the direct effect.

The simulation study is based on a data-generating process that partly mimics the settings of the HPTN 037 study (Latkin et al., 2009). We generated $n_e = 200$ disjoint ego-networks with two alters each, which was the average number of alters per ego-network in HPTN 037, resulting in $n_a = 400$. Contamination between the ego-networks was generated by specifying the expected number of missing alter-ego and ego-ego edges, as in Examples 2 and 4, and drawing the latent edges independently. Specifically, we set the expected number of missing alter-ego edges to $m^a \in \{100, 200\}$, and ego-ego edges to $m^e \in \{150, 250\}$. Under each setup, a single network was generated and was fixed throughout the simulation replications. The potential outcomes were generated once according to the models

$$\begin{aligned} Y_i(z, f) &= -0.5 + 2z + 0.5f + zf + \mathbf{X}_i^T \boldsymbol{\beta}_e + \varepsilon_i, & \text{for } i \in \mathcal{R}_e, \\ Y_i(0, f) &= -0.5 + 2f + \mathbf{X}_i^T \boldsymbol{\beta}_a + \varepsilon_i, & \text{for } i \in \mathcal{R}_a, \end{aligned}$$

where $\varepsilon_i \sim N(0, 1)$ are independent noise terms. This data-generating process yields true effects of $DE = IE = 2$, and ratio of direct effects among exposed versus non-exposed egos of $\kappa = 1.5$ as defined in (7). The generation process of the three covariates $\mathbf{X}_i$ and the specification of $\boldsymbol{\beta}_e$ and $\boldsymbol{\beta}_a$ are provided in Appendix D. Treatments were assigned according to a Bernoulli design with $p_z = 0.5$. We performed 5×10^3 simulation replications by random draws of treatment assignments.

We report here the results when contamination was generated using heterogeneous edge probabilities as in Example 4 with Euclidean distance and $\gamma^e = \gamma^a = 1$. We compare the bias-corrected estimators under the specification that uses the correct edge probabilities (“Heterogeneous”), under the specification that uses a misspecified model with the same edge probabilities as in Example 2 (“Homogeneous”) and the uncorrected estimators $\widehat{IE}$ and $\widehat{DE}$ (“Naive”). We considered both the HT estimators and the augmented estimators using an outcome model estimated via linear regression with two-fold cross-fitting (as described in Section 3.3).

Table 1 summarizes the results for the indirect and direct effect estimation. As the level of contamination (m^a and m^e) increases, the naive estimators exhibit substantial bias and poor coverage. In contrast, all bias-corrected estimators exhibit negligible bias across all scenarios, illustrating that even a misspecified model for the edge probabilities (homogeneous specification) can effectively correct for the bias. For the non-augmented estimators, the variance estimates were conservative, particularly for the direct effect in high contamination settings (SD/SE = 0.574). This is comparable to the conservativeness of the variance estimators in other papers in the network interference literature (e.g., Aronow and Samii, 2017; Weinstein and Nevo, 2026). The augmented indirect effect estimators, on the other hand, yielded valid variance estimates and achieved nominal coverage rates. The augmented direct effect estimators tend to slightly overestimate the variance (SD/SE < 1), which is attributed to their conservative variance estimator (Section 3.2). Results

for homogeneous contamination settings are provided in Appendix D and exhibit similar patterns.

We also evaluated the impacts of misspecifying the value of κ or violating Assumption 4. We found that violations of Assumption 4, induced by heterogeneous treatment-exposure interactions, resulted in negligible bias for the corrected direct effect estimators, albeit with slight variance underestimation. The impact of κ misspecification depended on the direction of the misspecification compared to the true value. If κ is misspecified in the wrong direction (e.g., specifying $\kappa < 1$ when $\kappa > 1$), the bias-corrected estimator can exhibit severe bias and lower coverage than the naive estimator. Conversely, when κ is misspecified in magnitude but in the correct direction, the bias-corrected estimator outperforms the naive estimator. Moreover, the performance of the proposed bias-corrected estimators under binary potential outcomes was also assessed, yielding results consistent with those under the continuous potential outcome model. See Appendix D for additional details and results.

Table 1: Simulation results for indirect and direct effects under heterogeneous contamination. Bias is the empirical bias, Coverage is the empirical coverage of the 95% confidence intervals, and SD/SE is the ratio of the standard deviation of the estimates to the mean standard error estimates. Specification “Heterogeneous” uses the correct edge probabilities, “Homogeneous” uses the same probability for all edges, and “Naive” is the uncorrected estimator. Augmented indicates whether the estimators augmented with an outcome model are used. True effects are $IE = 2$ and $DE = 2$.

Estimand	Scenario	Specification	Augmented	Bias	Coverage	SD/SE
<i>IE</i>	$m^a = 100$	Heterogeneous	FALSE	-0.006	0.963	0.948
			TRUE	-0.007	0.948	0.988
		Homogeneous	FALSE	-0.004	0.963	0.949
			TRUE	-0.005	0.949	0.987
		Naive	FALSE	-0.239	0.507	0.949
			TRUE	-0.240	0.403	0.987
	$m^a = 200$	Heterogeneous	FALSE	0.002	0.970	0.902
			TRUE	0.000	0.955	0.988
		Homogeneous	FALSE	0.011	0.971	0.904
			TRUE	0.009	0.954	0.987
		Naive	FALSE	-0.434	0.074	0.904
			TRUE	-0.435	0.026	0.987
<i>DE</i>	$m^e = 150, \kappa = 1.5$	Heterogeneous	FALSE	0.000	0.986	0.780
			TRUE	-0.005	0.960	0.952
		Homogeneous	FALSE	-0.011	0.984	0.802
			TRUE	-0.016	0.956	0.960
		Naive	FALSE	0.514	0.279	1.068
			TRUE	0.508	0.175	1.045
	$m^e = 250, \kappa = 1.5$	Heterogeneous	FALSE	0.000	0.999	0.574
			TRUE	-0.005	0.981	0.829
		Homogeneous	FALSE	-0.019	0.998	0.596
			TRUE	-0.024	0.976	0.840
		Naive	FALSE	0.689	0.102	1.064
			TRUE	0.682	0.023	1.029

5 Data analysis

HPTN 037 was an ENRT that evaluated the efficacy of a peer education intervention in reducing HIV risk behaviors among people who inject drugs and their drug and sexual networks (Latkin et al., 2009). Egos were randomized ($p_z = 0.5$) to receive a peer education training and were encouraged to share the HIV risk reduction information with their drug and sexual network members (alters). Egos in the control arm received standard HIV prevention counseling. The study was conducted between 2002 and 2006 in Philadelphia, Pennsylvania, US and Chiang Mai, Thailand. The study recruited 414 egos (index participants) who were asked to recruit at least one drug or sex network member (alter) into the study. Researchers enforced that alters were not also egos, and that each alter was connected to only one ego in the observed network (Latkin et al., 2009). Following previous analyses (Buchanan et al., 2018, 2024; Chao et al., 2025), we limit our analysis to the Pennsylvania site and to a 12-month follow-up, yielding a sample of $n_e = 150$ egos and $n_a = 263$ alters, with an average of 2 alters per ego-network. The primary outcome is a binary indicator of any injection-related risk behavior in the past month, including sharing injection equipment, front and back loading, and injecting with people not well known or in a public space.

Contamination across ego-networks is a concern in ENRT. That is, alters might inject drugs or have sexual relations with egos from other ego-networks, or egos might be connected to each other. Using recall data collected after the trial, Simmons et al. (2015) found evidence of contamination. Chao et al. (2025) developed a bias-correction method to adjust for (homogeneous) exposure misclassification, only for the indirect effect, building on the recall data and the assumption of non-overlapping ego-networks. That is, as in most ENRT studies, the analysis in Chao et al. (2025) assumed that, in the population network, alters can only be connected to one ego, and egos can not be connected to each other.

We relax this assumption and apply our proposed sensitivity analysis methods to assess the potential impact of contamination on both indirect and direct effect estimates. For both alter-ego and ego-ego edges, we specify the expected number of missing edges as in Examples 2 and 4. Covariates used in the analysis include risk behaviors at baseline, age, gender, and race. In the outcome models, we also included the average neighbors' covariates for each unit, as was also done by Buchanan et al. (2024). Quantitative covariates were scaled to have a mean of zero and a variance of one. For the specification of heterogeneous edge probabilities using the approximate expected number of missing edges (Example 4), we use the Euclidean distance and set $\gamma^e = \gamma^a = 1$. Results with $\gamma^e = \gamma^a = 2$ are given in Appendix E and were consistent with the results shown here, albeit with higher variance estimates. Note that here γ^e and γ^a are hyperparameters and not the sensitivity parameters, which are κ , m^a , and m^e . For the direct effect, we specify values larger than 1 to the sensitivity parameter κ , addressing the concern that the direct effect is larger, on average, among egos exposed to at least one other treated ego. We apply both the GSA and PBA described in Section 3.5. Variance estimates and 95% confidence intervals were calculated as described in Sections 3.1-3.3.

Furthermore, we compare the results of the sensitivity analysis with those obtained using point estimates of parameters from internal validation data, as described in Section 3.6. During the 6-month follow-up of the study, participants were asked to recall terms introduced in the intervention training program and unlikely to be known outside of it. Following previous analyses (Aroke et al., 2023; Chao et al., 2025), we assume that a participant was indirectly exposed to the intervention if she recalled at least one of the five terms associated with the intervention. For greater accuracy, we limit the validation subset only to partic-

ipants who recalled the positive term and none of the negative terms (Aroke et al., 2023; Chao et al., 2025). As described in Section 3.6, we estimate π^a from non-exposed alters ($\tilde{F}_i = 0, i \in \mathcal{R}_a$) and π^e from non-treated egos ($Z_i = 0, i \in \mathcal{R}_e$), both under homogeneous exposure probabilities. Among 22 alters and 15 egos in the validation subset, only 4 alters and 3 egos recalled the terms, mapping to an estimated exposure probabilities of $\hat{\pi}^a = 0.59$ and $\hat{\pi}^e = 0.2$. From Example 2 this is equivalent to approximately $\hat{m}^a = 105.5$ missing ego-alter edges and $\hat{m}^e = 33.4$ missing ego-ego edges.

5.1 Grid sensitivity analysis

For the indirect effect, we consider a grid of the approximate expected number of missing edges with $m^a \in [10, 500]$ and $m^e \in [10, 150]$, representing the concern that, in the homogeneous case, each alter and ego are approximately connected to at most two additional egos (Example 2). In addition, we consider $\kappa \in [1, 2]$ reflecting the suspicion that the direct effect among egos connected to at least one treated ego is up to two times larger, on average, than the effect among egos that are not connected to other treated egos. We report here the results for the augmented estimators $\widehat{IE}_{adj}^{aug}$ and $\widehat{DE}_{adj}^{aug}$ using two-fold cross-fitting (Section 3.3) and logistic regression for the outcome models. We consider both homogeneous and heterogeneous edge probability specifications. Results for the non-augmented estimators $\widehat{IE}_{adj}$ and $\widehat{DE}_{adj}$ are provided in Appendix E, and were consistent with the results shown here.

Figures 2 and 3 summarize the GSA results for indirect and direct effects, respectively. For the indirect effect, both homogeneous and heterogeneous specifications yield similar results. The results suggest that, for instance, if each alter is connected to up to one additional ego on average ($m^a = 263$) in the homogeneous case, the bias-corrected estimate is around -0.194 compared to the naive estimate of -0.076 , and a bias-corrected estimate of around -0.093 using the estimate $\hat{m}^a = 105.5$ from the internal validation data. Taking the estimate $\hat{m}^a = 105.5$ as a lower bound for the true m^a , implies that the naive estimator underestimate IE by an order of at least 18%, indicating a substantially larger indirect effect than previously reported.

For the direct effect, the bias-corrected estimates under the homogeneous and heterogeneous specifications yielded similar estimates that are closer to zero than the naive estimate of -0.116 across the entire grid of sensitivity parameters. For example, if $\kappa = 1.5$ and each ego is connected to up to one additional ego on average ($m^e = 75$), the bias-corrected estimate is around -0.097 in the homogeneous case, compared to the bias-corrected estimate of around -0.105 using the estimate $\hat{m}^e = 33.4$ from the internal validation data. These results imply that the naive estimates of DE might be overestimated by over 10.5%, indicating a smaller direct effect than previously found.

5.2 Probabilistic bias analysis

For the PBA, we specify distributions for the sensitivity parameters, representing different degrees of uncertainty about their values. For m^a and m^e , we compare three uncertainty distributions. The first is a discrete uniform distribution over the ranges used in the GSA: $m^a \sim \text{Uniform}\{1, 500\}$ and $m^e \sim \text{Uniform}\{1, 150\}$, representing maximum entropy, but with a capped right tail. The second is a Poisson distribution with a mean equal to half of the maximum value in the GSA grid: $m^a \sim \text{Poisson}(250)$ and $m^e \sim \text{Poisson}(75)$, representing moderate uncertainty with a thin but unbounded tail. The third is a Negative Binomial (NB) distribution with a mean equal to half the maximum value in the GSA grid and a size

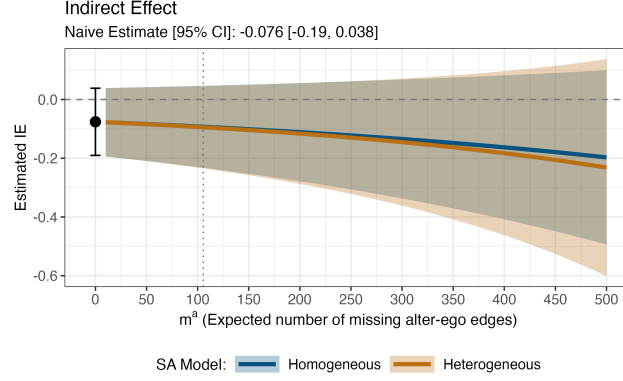


Figure 2: Grid sensitivity analysis results for indirect effect using the augmented estimator $\widehat{IE}_{adj}^{aug}$. The x-axis represents the approximate total number of missing alter-ego edges m^a , while the y-axis represents the estimates. The dashed vertical line at 105.5 is the estimate of m^a from the internal validation data. Results are shown for both homogeneous (blue) and heterogeneous (orange) edge probabilities specifications.

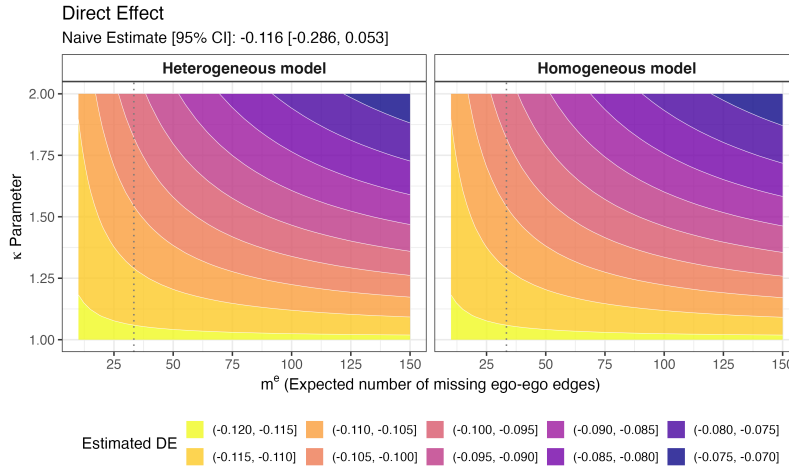


Figure 3: Grid sensitivity analysis results for direct effect using the augmented estimator $\widehat{DE}_{adj}^{aug}$. The x-axis represents the approximate total number of missing ego-ego edges m^e , while the y-axis represents κ , the interaction sensitivity parameter. The dashed vertical line at 33.4 is the estimate of m^e from the internal validation data. Contour lines and color shading represent the estimated values. Results are shown for both heterogeneous (left) and homogeneous (right) edge probability specifications.

parameter equal to 10, representing a heavier tail than the Poisson distribution. For the sensitivity parameter κ , we consider two uncertainty distributions. The first is a continuous uniform distribution, $\kappa \sim \text{Uniform}[1, 2]$, and the second is a log-normal distribution with expected value 1.5 and standard deviation equal to 0.2 on the log scale, resulting in a heavier right tail than the uniform distribution. We run the PBA procedure with $B = 10^4$ Monte Carlo samples and account for both bias and statistical uncertainty, as described in Section 3.5.

Figure 4 summarizes the PBA results for the indirect and direct effects with the augmented estimators. For the indirect effect, the three distributions for m^a yield similar average estimates, with a slightly larger magnitude for the heterogeneous specification. The uniform distribution results in the widest intervals, while the Poisson distribution yields the narrowest intervals. All bias-corrected estimates are lower (average estimates around -0.158) than the naive estimate of -0.076 , suggesting that contamination leads to severe underestimation of the magnitude of the indirect effect. For the direct effect, all specified distributions for m^e and κ yield similar average estimates, with the Poisson and uniform combination resulting in estimates closer to zero than the rest. The uniform distribution yields the widest intervals. The averages of the bias-corrected estimates are closer to zero (average estimates around -0.12 in the heterogeneous specification) than the naive estimate of -0.116 , suggesting that contamination leads to slight overestimation of the direct effect. Nevertheless, apart from the heterogeneous specification with the Poisson distribution for indirect effect, the indirect and direct 95% PBA intervals include zero in all distribution and edge probabilities specifications.

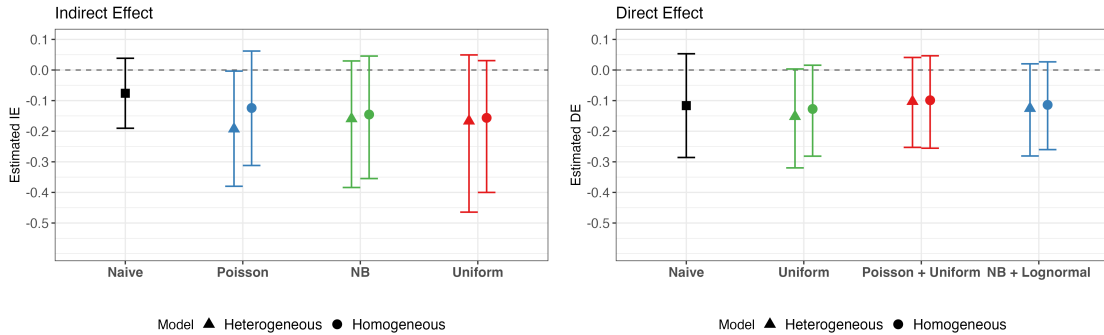


Figure 4: Probabilistic bias analysis results for indirect effect (left) using the augmented estimator $\widehat{IE}_{adj}^{aug}$, and direct effect (right) using the augmented estimator $\widehat{DE}_{adj}^{aug}$. Results are shown as mean and 95% intervals for both homogeneous and heterogeneous edge probability specifications, accounting for both bias and statistical uncertainty. Colors represent different specified distributions for the total number of missing edges and the sensitivity parameter κ .

6 Discussion

In this paper, we developed sensitivity analysis methods to assess the impact of contamination between ego-networks on the causal estimates in egocentric network randomized trials. Such trials are increasingly used to evaluate interventions in hard-to-reach populations, where full network information is often unavailable or difficult to obtain. Our methods build on a general formulation of contamination as exposure misclassification due to missing network edges, and provide bias-corrected estimators for both indirect and direct effects. Our bias-corrected estimators rely on two types of sensitivity parameters. The

first characterizes the magnitude of contamination. The second, specific to the direct effect, represents the ratio of the average direct effect among egos exposed to at least one treated neighbor to those not exposed. We provided practical guidelines for specifying these sensitivity parameters and described two sensitivity analysis approaches: GSA and PBA.

Our re-analysis of the HPTN 037 trial underscores the consequences of unmeasured contamination in ENRTs. Our findings demonstrate that standard estimators can severely underestimate the indirect effect while slightly overestimating the direct effect. For policy-makers, this highlights that peer-driven interventions may be substantially more effective than naive estimates suggest, altering the cost-benefit calculus for deploying network-based public health strategies.

A limitation of our design-based framework is the focus on sample-level causal estimands (1) and (2), which restrict our inference to the finite sample of recruited egos and alters rather than the broader population. Extensions to population-level estimands require specifying the exact mechanism of the egocentric network sampling, such as the inclusion probabilities of the egos. However, many ENRTs, including the HPTN 037 study, involve hard-to-reach populations where egos are typically recruited via convenience sampling rather than a strict probability sample. While recent methodological work has considered population-level estimands in ENRTs, doing so generally relies on the assumption of simple random sampling of units from the population (Fang et al., 2025). Consequently, our focus on sample-level estimands reflects a methodological trade-off: we prioritize robust internal validity and assumption-lean causal inference within the sample, without imposing strong unverifiable conditions on the recruitment process.

A natural alternative to our sensitivity framework might be to explicitly infer the missing alter-ego and ego-ego edges by assuming a stochastic generative model for the population network $\mathbf{A}$ and estimating its parameters from the observed network $\hat{\mathbf{A}}$. However, this approach presents significant challenges. First, it requires both positing a valid population-level network model and specifying the egocentric sampling mechanism. Second, even if these mechanisms could be specified, a fundamental challenge remains. Fitting a model to a network of n nodes does not necessarily recover the parameters governing the population network of N nodes. Assuming that edge formation probabilities scale directly from a sample to the population requires the network model to satisfy properties such as consistency under sampling (Shalizi and Rinaldo, 2013; Crane and Dempsey, 2021). Our design-based framework coupled with sensitivity analysis offers a reliable alternative. By treating the missing edges as fixed but uncertain, we can quantify the impact of contamination while bypassing the need to posit assumptions on the data-generating and sampling processes.

We assumed a two-level exposure mapping specification, indicating whether a unit is connected to at least one treated neighbor (Assumption 2). This specification is sensible when the treatment effect is saturated at one treated neighbor (regardless of who it is). That is, when exposure to additional treated neighbors does not further affect the outcome. In Appendix B, we extend our methods to a three-level exposure mapping specification, distinguishing between no treated neighbors, one treated neighbor, and two-or-more treated neighbors. In this case, for example, an alter with an observed exposure level of one treated neighbor may have a true exposure level of one or two-or-more treated neighbors. We limit our analysis for the indirect effect IE , and derive bias expressions for the naive estimator $\widehat{IE}$. We also provide bias-corrected estimators that depend on an additional sensitivity parameter that captures the marginal unit-level exposure effect. We show that the two-level exposure mapping model is a special case of the three-level model with saturation of effect after the first exposure.

The exposure mapping specification we took in this paper relies on a first-order neighbor-

hood interference assumption, restricting the diffusion of treatment effects to direct neighbors. A valuable direction for future research is relaxing this specification to accommodate higher-order interference. For example, under second-order interference, an alter could be indirectly affected by the treatment through a connection to another alter in a treated ego-network. In such settings, an ego might also have indirect exposure if its recruited alters are connected to other treated egos. Adapting our proposed sensitivity analysis framework to these scenarios would require specifying sensitivity parameters for latent alter-alter edges across different ego-networks. While conceptually natural, this extension would introduce notable combinatorial complexity to the computation of the exposure probabilities.

ENRT is an attractive design that allows for the estimation of spillovers and the indirect effect without the requirement for sociocentric network measurements. It enables researchers to estimate causal effects under interference with limited network data. However, extending these designs to observational studies where units are sampled from a larger network (e.g., through egocentric sampling), but treatments are not randomly assigned by the researchers, is a more challenging problem. The primary concern is that non-recruited units can be treated as well, leading to treatment interference between recruited and non-recruited units. Since information on the edges connecting such units in the population is rarely available, addressing interference in observational studies with egocentric sampling remains a significant challenge.

Funding

This work was supported by the Israel Science Foundation (ISF grant No. 2300/25); BW is supported by the Data Science Fellowship granted by the Israeli Council for Higher Education.

Acknowledgements

We thank Dr. Ashley Buchanan for sharing code and assisting in the analysis of the HPTN 037 study. We also thank two anonymous reviewers for helpful comments and suggestions that improved the paper.

Conflict of interest

None declared.

Data Availability

Analysis code applied to a simulated dataset similar to the design of HPTN 037 can be found at https://github.com/barwein/ENRT_SA. The HPTN 037 study datasets are publicly available and can be requested from the Statistical Center for HIV/AIDS Research and Prevention through <https://atlas.scharp.org/project/home/begin.view>.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296.

- Aroke, H., Buchanan, A., Katenka, N., Crawford, F. W., Lee, T., Halloran, M. E., and Latkin, C. (2023). Evaluating the mediating role of recall of intervention knowledge in the relationship between a peer-driven intervention and hiv risk behaviors among people who inject drugs. *AIDS and Behavior*, 27(2):578–590.
- Aronow, P. M. and Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4).
- Asmus, J. M., Carter, E. W., Moss, C. K., Biggs, E. E., Bolt, D. M., Born, T. L., Bottema-Beutel, K., Brock, M. E., Cattey, G. N., Cooney, M., Fesperman, E. S., Hochman, J. M., Huber, H. B., Lequia, J. L., Lyons, G. L., Vincent, L. B., and Weir, K. (2017). Efficacy and Social Validity of Peer Network Interventions for High School Students With Severe Disabilities. *American Journal on Intellectual and Developmental Disabilities*, 122(2):118–137.
- Baldi, P. and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *The Annals of Probability*, 17(4):1646–1650.
- Booth, R. E., Davis, J. M., Dvoryak, S., Brewster, J. T., Lisovska, O., Strathdee, S. A., and Latkin, C. A. (2016). HIV incidence among people who inject drugs (PWIDs) in Ukraine: results from a clustered randomised trial. *The Lancet HIV*, 3(10):e482–e489.
- Boucher, V. and Houndetoungan, E. A. (2022). Estimating peer effects using partial network data. *Centre de recherche sur les risques les enjeux économiques et les politiques*.
- Buchanan, A. L., Hernández-Ramírez, R. U., Lok, J. J., Vermund, S. H., Friedman, S. R., Forastiere, L., and Spiegelman, D. (2024). Assessing direct and spillover effects of intervention packages in network-randomized studies. *Epidemiology*, 35(4):481.
- Buchanan, A. L., Vermund, S. H., Friedman, S. R., and Spiegelman, D. (2018). Assessing Individual and Disseminated Effects in Network-Randomized Studies. *American Journal of Epidemiology*, 187(11):2449–2459.
- Chandrasekhar, A. and Lewis, R. (2011). Econometrics of sampled networks. *Unpublished manuscript, MIT*. [422].
- Chao, A., Spiegelman, D., Buchanan, A., and Forastiere, L. (2025). Estimation and inference for causal spillover effects in egocentric-network randomized trials in the presence of network membership misclassification. *Biostatistics*, 26(1):kxaf009.
- Cox, D. R. (1958). *Planning of experiments*. Wiley, New York.
- Crane, H. and Dempsey, W. (2021). A Statistical Framework for Modern Network Science. *Statistical Science*, 36(1).
- Crawford, F. W., Morozova, O., Buchanan, A. L., and Spiegelman, D. (2019). Interpretation of the Individual Effect Under Treatment Spillover. *American Journal of Epidemiology*, 188(8):1407–1409.
- Desrosiers, A., Kumar, P., Dayal, A., Alex, L., Akram, A., and Betancourt, T. (2020). Diffusion and spillover effects of an evidence-based mental health intervention among peers and caregivers of high risk youth in Sierra Leone: study protocol. *BMC Psychiatry*, 20(1):85.

- Ding, P. (2024). *A first course in causal inference*. Chapman and Hall/CRC.
- Egami, N. (2020). Spillover effects in the presence of unobserved networks. *Political Analysis*, 29(3):287–316.
- Fang, J., Spiegelman, D., Buchanan, A. L., and Forastiere, L. (2025). Design of egocentric network-based studies to estimate causal effects under interference. *Statistical Methods in Medical Research*, page 09622802251357021.
- Forastiere, L., Airoidi, E. M., and Mealli, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918.
- Fox, M. P., MacLehose, R. F., and Lash, T. L. (2021). *Applying quantitative bias analysis to epidemiologic data*, volume 10. Springer.
- Gao, M. and Ding, P. (2025). Causal inference in network experiments: regression-based analysis and design-based properties. *Journal of Econometrics*, 252:106119.
- Greenland, S. (1996). Basic methods for sensitivity analysis of biases. *International journal of epidemiology*, 25(6):1107–1116.
- Greenland, S. (2005). Multiple-Bias Modelling for Analysis of Observational Data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 168(2):267–306.
- Halloran, M. E. and Struchiner, C. J. (1991). Study designs for dependent happenings. *Epidemiology*, 2(5):331–338.
- Hardy, M., Heath, R. M., Lee, W., and McCormick, T. H. (2019). Estimating spillovers using imprecisely measured networks. *arXiv preprint arXiv:1904.00136*.
- Hayes, R. J. and Moulton, L. H. (2017). *Cluster Randomised Trials*. CRC PR INC.
- Hong, Y. (2013). On computing the distribution function for the poisson binomial distribution. *Computational Statistics & Data Analysis*, 59:41–51.
- Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.
- Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.
- Lash, T. L., Fox, M. P., MacLehose, R. F., Maldonado, G., McCandless, L. C., and Greenland, S. (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology*, 43(6):1969–1985.
- Latkin, C., Donnell, D., Liu, T.-Y., Davey-Rothwell, M., Celentano, D., and Metzger, D. (2013). The dynamic relationship between social norms and behaviors: the results of an HIV prevention network intervention for injection drug users. *Addiction*, 108(5):934–943.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/add.12095>.
- Latkin, C. A., Donnell, D., Metzger, D., Sherman, S., Aramrattana, A., Davis-Vogel, A., Quan, V. M., Gandham, S., Vongchak, T., Perdue, T., and Celentano, D. D. (2009). The efficacy of a network intervention to reduce HIV risk behaviors among drug users and risk partners in Chiang Mai, Thailand and Philadelphia, USA. *Social Science & Medicine*, 68(4):740–748.

- Leung, M. P. (2025). Cluster-randomized trials with cross-cluster interference. *Journal of the American Statistical Association*, pages 1–11.
- Li, W., Sussman, D. L., and Kolaczyk, E. D. (2021). Causal inference under network interference with noise. *arXiv preprint arXiv:2105.04518*.
- Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman’s critique. *The Annals of Applied Statistics*, pages 295–318.
- Lu, X., Shi, L., Liu, H., and Ding, P. (2025). Conditional cross-fitting for unbiased machine-learning-assisted covariate adjustment in randomized experiments. *arXiv preprint arXiv:2508.15664*.
- Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23.
- Marray, K. (2024). Estimating spillovers from sampled connections. *arXiv preprint arXiv:2410.17154*.
- Marsden, P. V. (2002). Egocentric and sociocentric measures of network centrality. *Social networks*, 24(4):407–422.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.
- Miller, W. C., Hoffman, I. F., Hanscom, B. S., Ha, T. V., Dumchev, K., Djoerban, Z., Rose, S. M., Latkin, C. A., Metzger, D. S., Lancaster, K. E., Go, V. F., Dvoriak, S., Mollan, K. R., Reifeis, S. A., Piwowar-Manning, E. M., Richardson, P., Hudgens, M. G., Hamilton, E. L., Sugarman, J., Eshleman, S. H., Susami, H., Chu, V. A., Djauzi, S., Kiriazova, T., Bui, D. D., Strathdee, S. A., and Burns, D. N. (2018). A scalable, integrated intervention to engage people who inject drugs in HIV care and medication-assisted treatment (HPTN 074): a randomised, controlled phase 3 feasibility and efficacy study. *The Lancet*, 392(10149):747–759. Publisher: Elsevier.
- Ogburn, E. L., Sofrygin, O., Díaz, I., and van der Laan, M. J. (2024). Causal inference for social network data. *Journal of the American Statistical Association*, pages 1–46.
- Perry, B. L., Pescosolido, B. A., and Borgatti, S. P. (2018). *Egocentric Network Analysis: Foundations, Methods, and Models*. Structural Analysis in the Social Sciences. Cambridge University Press, Cambridge.
- Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218.
- Sävje, F. (2024). Causal inference with misspecified exposure mappings: separating definitions and assumptions. *Biometrika*, 111(1):1–15.
- Sewell, D. K. (2017). Network autocorrelation models with egocentric data. *Social Networks*, 49:113–123.
- Shalizi, C. R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics*, 41(2):508–535. Publisher: Institute of Mathematical Statistics.

- Sherman, S. G., Sutcliffe, C., Sirojn, B., Latkin, C. A., Aramratanna, A., and Celentano, D. D. (2009). Evaluation of a peer network intervention trial among young methamphetamine users in Chiang Mai, Thailand. *Social Science & Medicine*, 68(1):69–79.
- Simmons, N., Donnell, D., Ou, S.-s., Celentano, D. D., Aramrattana, A., Davis-Vogel, A., Metzger, D., and Latkin, C. (2015). Assessment of Contamination and Misclassification Biases in a Randomized Controlled Trial of a Social Network Peer Education Intervention to Reduce HIV risk Behaviors Among Drug Users and Risk Partners in Philadelphia, PA and Chiang Mai, Thailand. *AIDS and Behavior*, 19(10):1818–1827.
- Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling (Springer Series in Statistics)*. Springer.
- Tchetgen Tchetgen, E. J., Fulcher, I. R., and Shpitser, I. (2020). Auto-g-computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844.
- Toulis, P. and Kao, E. (2013). Estimation of causal peer influence effects. In *International conference on machine learning*, pages 1489–1497. PMLR.
- VanderWeele, T. J., Tchetgen, E. J. T., and Halloran, M. E. (2014). Interference and sensitivity analysis. *Statistical Science*, 29(4).
- Weinstein, B. and Nevo, D. (2025). Bayesian estimation of causal effects using proxies of a latent interference network. *arXiv preprint arXiv:2505.08395*.
- Weinstein, B. and Nevo, D. (2026). Causal inference with misspecified network interference structure. *Biometrics*, 82(1):ujag023.
- Yauck, M. (2022). On the estimation of peer effects for sampled networks. *arXiv preprint arXiv:2208.09102*.

Appendix – Table of Contents

A	Proofs and technical details	30
A.1	Proof of Proposition 1	30
A.2	Exposure probabilities of alters	33
A.3	Proof of Proposition 2	33
A.4	Proof of Proposition 3	34
A.5	Direction of bias relative to the null	35
A.6	Direction and magnitude of bias-corrected estimators	36
A.7	Variance estimation and asymptotic distribution (no cross-fitting)	36
A.8	Augmented estimation via cross-fitting	39
A.9	Relative risk estimand and estimators	41
A.10	Relaxing Assumption 4	42
A.11	Estimating the sensitivity parameters with internal validation data	44
B	Extension to three-level exposure mapping	46
B.1	Preliminaries	46
B.2	Bias of the naive estimator	47
B.3	Bias-corrected estimator	49
B.4	Relation to the two-level exposure model	50
B.5	Direction of bias, sign preservation, and effect magnitude	50
B.6	Sensitivity analysis	50
C	Probabilistic bias analysis	51
D	Simulations	52
E	Data analysis	55
F	Results of the example analysis code	58

A Proofs and technical details

A.1 Proof of Proposition 1

Recall that the HT estimator of the indirect effect based on the observed data (3) is

$$\widehat{IE} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right]$$

By the assumed experimental design (Assumption 1), the probability that an alter is assigned to treatment is zero, that is, $\Pr(Z_i = 1 \mid i \in \mathcal{R}_a) = 0$. In addition, by design Z_i and $\tilde{F}_i = F(\mathbf{Z}_{-i}, \tilde{\mathbf{A}})$ are independent. Thus, for an alter $i \in \mathcal{R}_a$, we have $\mathbb{I}\{\tilde{F}_i = f\} = \mathbb{I}\{Z_i = 0, \tilde{F}_i = f\}$ with probability one. Starting with the first inner term, for

$i \in \mathcal{R}_a$,

$$\begin{aligned}
& \mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} \right] \\
& \stackrel{(i)}{=} \frac{1}{p_z} \mathbb{E}_{\mathbf{Z}} \left[\mathbb{I}\{F_i = 1, \tilde{F}_i = 1\} Y_i(0, 1) + \mathbb{I}\{F_i = 0, \tilde{F}_i = 1\} Y_i(0, 0) \right] \\
& \stackrel{(ii)}{=} \frac{\Pr(F_i = 1, \tilde{F}_i = 1 \mid i \in \mathcal{R}_a)}{p_z} Y_i(0, 1) + \frac{\Pr(F_i = 0, \tilde{F}_i = 1 \mid i \in \mathcal{R}_a)}{p_z} Y_i(0, 0) \tag{A.1} \\
& \stackrel{(iii)}{=} \Pr(F_i = 1 \mid \tilde{F}_i = 1, i \in \mathcal{R}_a) Y_i(0, 1) + \Pr(F_i = 0 \mid \tilde{F}_i = 1, i \in \mathcal{R}_a) Y_i(0, 0) \\
& \stackrel{(iv)}{=} Y_i(0, 1),
\end{aligned}$$

where (i) follows since $Z_i = 0$ with probability one for alters and consistency (Assumption 3) that connects the observed outcomes Y_i to the potential outcomes $Y_i(z, f)$; (ii) from taking expectation with respect to the design $\mathbf{Z}$ (recalling that the distribution of $\mathbf{Z}$ in this paper is conditional on the sample); (iii) simply uses that p_z is equal to $\Pr(\tilde{F}_i = 1 \mid i \in \mathcal{R}_a)$ and $\frac{\Pr(F_i=f, \tilde{F}_i=1 \mid i \in \mathcal{R}_a)}{\Pr(\tilde{F}_i=1 \mid i \in \mathcal{R}_a)} = \Pr(F_i = f \mid \tilde{F}_i = 1, i \in \mathcal{R}_a)$; and (iv) follows since $\Pr(F_i = 1 \mid \tilde{F}_i = 1, i \in \mathcal{R}_a) = 1$. To see this, note that $\sum_{j \neq i} Z_j \tilde{A}_{ij} \leq \sum_{j \neq i} Z_j A_{ij}$ with probability one, which implies that $\tilde{F}_i \leq F_i$ with probability one as well. Since F_i is binary, if $\tilde{F}_i = 1$, it must also be $F_i = 1$.

Moving to the second inner term, a similar derivation yields

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right] &= \Pr(F_i = 1 \mid \tilde{F}_i = 0, i \in \mathcal{R}_a) Y_i(0, 1) \\
&\quad + \Pr(F_i = 0 \mid \tilde{F}_i = 0, i \in \mathcal{R}_a) Y_i(0, 0).
\end{aligned}$$

By Bayes' rule,

$$\begin{aligned}
\Pr(F_i = 0 \mid \tilde{F}_i = 0, i \in \mathcal{R}_a) &= \frac{\Pr(\tilde{F}_i = 0 \mid F_i = 0, i \in \mathcal{R}_a) \Pr(F_i = 0 \mid i \in \mathcal{R}_a)}{\Pr(\tilde{F}_i = 0 \mid i \in \mathcal{R}_a)} \\
&= \frac{\Pr(F_i = 0 \mid i \in \mathcal{R}_a)}{\Pr(\tilde{F}_i = 0 \mid i \in \mathcal{R}_a)} \tag{A.2} \\
&= \frac{1 - \pi_i^a}{1 - p_z},
\end{aligned}$$

as $\Pr(\tilde{F}_i = 0 \mid F_i = 0, i \in \mathcal{R}_a) = 1$ since $\tilde{F}_i \leq F_i$ with probability one. Setting $\Pr(\tilde{F}_i = 1 \mid F_i = 0, i \in \mathcal{R}_a) = 1 - \Pr(\tilde{F}_i = 0 \mid F_i = 0, i \in \mathcal{R}_a)$ and rearranging terms yields

$$\mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right] = \frac{(\pi_i^a - p_z) Y_i(0, 1) + (1 - \pi_i^a) Y_i(0, 0)}{1 - p_z}. \tag{A.3}$$

Consequently, combining (A.1) and (A.3),

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}} [\widehat{IE}] &= \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \left[Y_i(0, 1) - \frac{(\pi_i^a - p_z) Y_i(0, 1) + (1 - \pi_i^a) Y_i(0, 0)}{1 - p_z} \right] \\
&= \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \frac{1 - \pi_i^a}{1 - p_z} [Y_i(0, 1) - Y_i(0, 0)] \\
&= IE + \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \frac{p_z - \pi_i^a}{1 - p_z} [Y_i(0, 1) - Y_i(0, 0)],
\end{aligned} \tag{A.4}$$

which completes the proof of the first part of Proposition 1.

Moving to the direct effect estimator. Recall that the HT estimator based on the observed data (4) is

$$\widehat{DE} = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \left[\frac{\mathbb{I}\{Z_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{Z_i = 0\} Y_i}{1 - p_z} \right].$$

By the assumed experimental design (Assumption 1), the observed exposure of all egos is zero $\Pr(\tilde{F}_i = 0 \mid i \in \mathcal{R}_e) = 1$, and among the egos $i \in \mathcal{R}_e$, we have that $\mathbb{I}\{Z_i = z\} = \mathbb{I}\{Z_i = z, \tilde{F}_i = 0\}$ with probability one. The expectation of the first inner term of $\widehat{DE}$, for $i \in \mathcal{R}_e$, is as follows.

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{Z_i = 1\} Y_i}{p_z} \right] &\stackrel{(i)}{=} \mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{Z_i = 1, \tilde{F}_i = 0\} Y_i}{p_z} \right] \\
&\stackrel{(ii)}{=} \frac{1}{p_z} \mathbb{E}_{\mathbf{Z}} \left[\mathbb{I}\{Z_i = 1, F_i = 1, \tilde{F}_i = 0\} Y_i(1, 1) \right] \\
&\quad + \frac{1}{p_z} \mathbb{E}_{\mathbf{Z}} \left[\mathbb{I}\{Z_i = 1, F_i = 0, \tilde{F}_i = 0\} Y_i(1, 0) \right] \\
&\stackrel{(iii)}{=} \Pr(F_i = 1, \tilde{F}_i = 0 \mid i \in \mathcal{R}_e) Y_i(1, 1) \\
&\quad + \Pr(F_i = 0, \tilde{F}_i = 0 \mid i \in \mathcal{R}_e) Y_i(1, 0) \\
&\stackrel{(iv)}{=} \Pr(F_i = 0 \mid i \in \mathcal{R}_e) Y_i(1, 1) \\
&\quad + \Pr(F_i = 1 \mid i \in \mathcal{R}_e) Y_i(1, 0),
\end{aligned}$$

where (i) follows from the explanation described above; (ii) from consistency (Assumption 3); (iii) since Z_i is independent of both F_i and $\tilde{F}_i$ by Assumption 1, and $\Pr(Z_i = 1 \mid i \in \mathcal{R}_e) = p_z$; and (iv) since $\tilde{F}_i$ is equal to 0 with probability one. Recalling that $\pi_i^e = \Pr(F_i = 1 \mid i \in \mathcal{R}_e)$ and plugging it into the above expectation yields

$$\mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{Z_i = 1\} Y_i}{p_z} \right] = \pi_i^e Y_i(1, 1) + (1 - \pi_i^e) Y_i(1, 0). \tag{A.5}$$

Replacing $Z_i = 1$ with $Z_i = 0$ and p_z with $1 - p_z$ above, yields similar results. Specifically, for $z = 0$ we have

$$\mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{Z_i = 0\} Y_i}{1 - p_z} \right] = \pi_i^e Y_i(0, 1) + (1 - \pi_i^e) Y_i(0, 0). \tag{A.6}$$

Plugging (A.5) and (A.6) into $\mathbb{E}_{\mathbf{Z}} [\widehat{DE}]$ yields

$$\begin{aligned}\mathbb{E}_{\mathbf{Z}} [\widehat{DE}] &= \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} [\pi_i^e \{Y_i(1, 1) - Y_i(0, 1)\} + (1 - \pi_i^e) \{Y_i(1, 0) - Y_i(0, 0)\}] \\ &= DE + \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \pi_i^e [\{Y_i(1, 1) - Y_i(0, 1)\} - \{Y_i(1, 0) - Y_i(0, 0)\}],\end{aligned}\tag{A.7}$$

which completes the second part of Proposition 1. $\square$

A.2 Exposure probabilities of alters

Alter $i \in \mathcal{R}_a$ is exposed ($F_i = 1$) if its ego $e(i)$ or at least one of the other recruited egos connected to i in the population network $\mathbf{A}$ is treated. We have

$$\begin{aligned}\pi_i^a &= \Pr(F_i = 1 \mid i \in \mathcal{R}_a) \\ &= \Pr(F_i = 1 \mid Z_{e(i)} = 1, i \in \mathcal{R}_a) \Pr(Z_{e(i)} = 1 \mid i \in \mathcal{R}_a) \\ &\quad + \Pr(F_i = 1 \mid Z_{e(i)} = 0, i \in \mathcal{R}_a) \Pr(Z_{e(i)} = 0 \mid i \in \mathcal{R}_a),\end{aligned}\tag{A.8}$$

but $\Pr(F_i = 1 \mid Z_{e(i)} = 1, i \in \mathcal{R}_a) = 1$ and $\Pr(Z_{e(i)} = 1 \mid i \in \mathcal{R}_a) = p_z \mathbb{I}\{e(i) \in \mathcal{R}_e\} = p_z$ by Assumptions 1 and 2. Since only recruited egos have a non-zero probability of being assigned to treatment, then for alter $i \in \mathcal{R}_a$, $F_i = \sum_{j \neq i} Z_j A_{ij} = \sum_{j \in \mathcal{R}_e} Z_j A_{ij}$ with probability one. Therefore,

$$\begin{aligned}\Pr(F_i = 0 \mid Z_{e(i)} = 0, i \in \mathcal{R}_a) &= \Pr\left(\sum_{j \in \mathcal{R}_e} Z_j A_{ij} = 0 \mid Z_{e(i)} = 0, i \in \mathcal{R}_a\right) \\ &= \Pr\left(\sum_{j \in \mathcal{R}_e \setminus \{e(i)\}} Z_j A_{ij} = 0 \mid i \in \mathcal{R}_a\right) \\ &= \prod_{j \in \mathcal{R}_e \setminus \{e(i)\}} (1 - p_z \rho_{ij}^a).\end{aligned}$$

Substituting $\Pr(F_i = 1 \mid Z_{e(i)} = 0, i \in \mathcal{R}_a) = 1 - \Pr(F_i = 0 \mid Z_{e(i)} = 0, i \in \mathcal{R}_a)$ in (A.8) yields (14).

A.3 Proof of Proposition 2

The bias-corrected estimator is

$$\widehat{IE}_{adj} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \frac{1 - p_z}{1 - \pi_i^a} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right].$$

As the weights $\frac{1 - p_z}{1 - \pi_i^a}$ are fixed and therefore independent of $\mathbf{Z}$, taking the expectation $\mathbb{E}_{\mathbf{Z}} [\widehat{IE}_{adj}]$ will result in the second row of (A.4) where the terms $\frac{1 - \pi_i^a}{1 - p_z}$ and $\frac{1 - p_z}{1 - \pi_i^a}$ are multiplied and therefore canceled. That implies $\mathbb{E}_{\mathbf{Z}} [\widehat{IE}_{adj}] = IE$ and completes the proof of Proposition 2. $\square$

A.4 Proof of Proposition 3

Recall that (7) states that

$$\overline{\Delta}(1) = \kappa \overline{\Delta}(0),$$

where $\overline{\Delta}(f) = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \Delta_i(f)$ and $\Delta_i(f) = Y_i(1, f) - Y_i(0, f)$. The bias-adjusted direct effect estimator is

$$\widehat{DE}_{adj} = \frac{1}{1 + \overline{\pi}^e(\kappa - 1)} \widehat{DE} = \frac{1}{u_e} \sum_{i \in \mathcal{R}_e} \left[\frac{\mathbb{I}\{Z_i = 1\}Y_i}{p_z} - \frac{\mathbb{I}\{Z_i = 0\}Y_i}{1 - p_z} \right],$$

where $u_e = n_e [1 + \overline{\pi}^e(\kappa - 1)]$ and $\overline{\pi}^e = n_e^{-1} \sum_{i \in \mathcal{R}_e} \pi_i^e$ is the average probability of exposure among the egos.

The inner product of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ can be written as $\frac{1}{m} \sum_i a_i b_i = \bar{a}\bar{b} + S_{\mathbf{a}, \mathbf{b}}$ where $S_{\mathbf{a}, \mathbf{b}} = \frac{1}{m} \sum_i (a_i - \bar{a})(b_i - \bar{b})$. Since both κ and π_i^e are fixed and therefore independent of $\mathbf{Z}$, we can use (A.7) and write the expectation of the naive estimator as

$$\begin{aligned} \mathbb{E}_{\mathbf{Z}} [\widehat{DE}] &= DE + \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \pi_i^e [\{Y_i(1, 1) - Y_i(0, 1)\} - \{Y_i(1, 0) - Y_i(0, 0)\}] \\ &= DE + \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \pi_i^e [\Delta_i(1) - \Delta_i(0)] \\ &= DE + \overline{\pi}^e (\overline{\Delta}(1) - \overline{\Delta}(0)) + S_{\pi^e, \Delta(1)} - S_{\pi^e, \Delta(0)} \\ &= [1 + \overline{\pi}^e(\kappa - 1)] DE + S_{\pi^e, \Delta(1)} - S_{\pi^e, \Delta(0)}, \end{aligned} \tag{A.9}$$

where we used $\overline{\Delta}(1) = \kappa \overline{\Delta}(0)$ and the fact that $\overline{\Delta}(0) = DE$. Thus, the expected value of the bias-corrected estimator is

$$\mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] = DE + \frac{S_{\pi^e, \Delta(1)} - S_{\pi^e, \Delta(0)}}{1 + \overline{\pi}^e(\kappa - 1)}.$$

Finally, under Assumption 4, $S_{\pi^e, \Delta(1)} = S_{\pi^e, \Delta(0)}$, and we obtain $\mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] = DE$. That completes the proof of Proposition 3. $\square$

In the case of homogeneous exposure probabilities $\pi_i^e = \pi^e$ for all $i \in \mathcal{R}_e$ for some constant $\pi^e \in [0, 1]$, we have $S_{\pi^e, \Delta(1)} = S_{\pi^e, \Delta(0)} = 0$, therefore $\widehat{DE}_{adj}$ is an unbiased estimator of DE .

Conditions under which the naive HT estimator is unbiased. If $\pi_i^e = 0$ for all $i \in \mathcal{R}_e$ or $\kappa = 1$, the bias-corrected estimator $\widehat{DE}_{adj}$ reduces to the naive HT estimator $\widehat{DE}$. If $\pi_i^e = 0$, then from Proposition 1 the naive HT estimator will be an unbiased estimator of DE . If $\kappa = 1$, we have two options

1. If the exposure probabilities are homogeneous ($\pi_i^e = \pi^e$), then $S_{\pi^e, \Delta(1)} = S_{\pi^e, \Delta(0)} = 0$ and $\widehat{DE}$ is an unbiased estimator of DE . That will be true even if there is some contamination between the egos, i.e., if $\pi^e > 0$.
2. As described in the main text, under the saturated potential outcome model

$$Y_i(z, f) = \beta_{0i} + \beta_{1i}z + \beta_{2i}f + \beta_{3i}zf,$$

Assumption 4 reduces to

$$S_{\pi^e, \beta_3} = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} (\pi_i^e - \bar{\pi}^e)(\beta_{3i} - \bar{\beta}_3) = 0.$$

Setting $\kappa = 1$ implies that $\sum_{i \in \mathcal{R}_e} \beta_{3i} = 0$. Thus, $\widehat{DE}$ will be an unbiased estimator of DE if $\sum_{i \in \mathcal{R}_e} \pi_i^e \beta_{3i} = 0$.

A.5 Direction of bias relative to the null

We now show the direction of bias of the naive estimators $\widehat{IE}$ and $\widehat{DE}$ relative to the null of no effects.

Indirect effect estimator Starting with $\widehat{IE}$. If $Y_i(0, 1) \geq Y_i(0, 0)$ for all $i \in \mathcal{R}_a$ such that $IE \geq 0$, then since $p_z - \pi_i^a \leq 0$, from (A.4), we have $0 \leq \mathbb{E}_{\mathbf{Z}}[\widehat{IE}] \leq IE$. Vice versa, if $Y_i(0, 1) \leq Y_i(0, 0)$ for all $i \in \mathcal{R}_a$, then $IE \leq 0$ and $0 \geq \mathbb{E}_{\mathbf{Z}}[\widehat{IE}] \geq IE$. That is, the estimator $\widehat{IE}$ is, in expectation, biased towards the null of no indirect effect. That will hold in both the homogeneous and heterogeneous scenarios of ρ_{ij}^a , as $p_z - \pi_i^a \leq 0$ for all $i \in \mathcal{R}_a$.

Direct effect estimator We show the direction of bias relative to the null of the naive direct effect estimator $\widehat{DE}$ using the representation κ as defined in (7). From (A.9)

$$\mathbb{E}_{\mathbf{Z}}[\widehat{DE}] = [1 + \bar{\pi}^e(\kappa - 1)] DE + S_{\pi^e, \Delta(1)} - S_{\pi^e, \Delta(0)}.$$

Under Assumption 4, the covariance terms are equal, thus

$$\mathbb{E}_{\mathbf{Z}}[\widehat{DE}] = [1 + \bar{\pi}^e(\kappa - 1)] DE$$

Since $\bar{\pi}^e \geq 0$, we have

- If $DE \geq 0$

$$\begin{cases} \mathbb{E}_{\mathbf{Z}}[\widehat{DE}] \geq DE \geq 0, & \kappa > 1, \\ DE \geq \mathbb{E}_{\mathbf{Z}}[\widehat{DE}] \geq 0, & \kappa \in (1 - \frac{1}{\bar{\pi}^e}, 1), \\ \mathbb{E}_{\mathbf{Z}}[\widehat{DE}] \leq 0, & \kappa < 1 - \frac{1}{\bar{\pi}^e}. \end{cases}$$

- If $DE \leq 0$

$$\begin{cases} \mathbb{E}_{\mathbf{Z}}[\widehat{DE}] \leq DE \leq 0, & \kappa > 1, \\ DE \leq \mathbb{E}_{\mathbf{Z}}[\widehat{DE}] \leq 0, & \kappa \in (1 - \frac{1}{\bar{\pi}^e}, 1), \\ \mathbb{E}_{\mathbf{Z}}[\widehat{DE}] \geq 0, & \kappa < 1 - \frac{1}{\bar{\pi}^e}. \end{cases}$$

Therefore, the estimator $\widehat{DE}$ is biased *away* from the null of no direct effect if $\kappa > 1$, biased *towards* the null if $\kappa \in (1 - \frac{1}{\bar{\pi}^e}, 1)$, and changes sign if $\kappa < 1 - \frac{1}{\bar{\pi}^e}$. Since π_i^e is typically small, then $1 - \frac{1}{\bar{\pi}^e}$ is a large negative number, and it is unlikely that κ is smaller than that value in practice. Thus, in general, we expect that the estimator $\widehat{DE}$ is biased away from the null if $\kappa > 1$ and towards the null if $\kappa < 1$ for many realistic scenarios.

A.6 Direction and magnitude of bias-corrected estimators

We now derive the direction and effect magnitude of the bias-corrected estimators $\widehat{IE}_{adj}$ and $\widehat{DE}_{adj}$ compared to their respective naive estimators $\widehat{IE}$ and $\widehat{DE}$.

Indirect effect estimators. Eq. (14) implies that $\pi_i^a \geq p_z$, therefore $\frac{1-p_z}{1-\pi_i^a} \geq 1$. If $\pi_i^a = \pi^a$ for all $i \in \mathcal{R}_a$, that is, the same for all alters, then $\widehat{IE}_{adj} = \frac{1-p_z}{1-\pi^a} \widehat{IE}$. Thus, $\text{sign}(\widehat{IE}_{adj}) = \text{sign}(\widehat{IE})$ and

$$\begin{cases} \widehat{IE}_{adj} \geq \widehat{IE}, & \widehat{IE} > 0, \\ \widehat{IE}_{adj} \leq \widehat{IE}, & \widehat{IE} < 0. \end{cases}$$

Direct effect estimators. Recall that $\widehat{DE}_{adj} = \frac{1}{1+\bar{\pi}^e(\kappa-1)} \widehat{DE}$. Thus, if $1 + \bar{\pi}^e(\kappa - 1) > 1 \Rightarrow \kappa > 1 - \frac{1}{\bar{\pi}^e}$, then $\text{sign}(\widehat{DE}_{adj}) = \text{sign}(\widehat{DE})$. Otherwise, if $\kappa < 1 - \frac{1}{\bar{\pi}^e}$, then $\text{sign}(\widehat{DE}_{adj}) = -\text{sign}(\widehat{DE})$. However, for example, if $\bar{\pi}^e \approx 0.01$, $1 - \frac{1}{\bar{\pi}^e} \approx -99$, and it is unlikely that κ is smaller than that value in practice. So in general, we expect $\text{sign}(\widehat{DE}_{adj}) = \text{sign}(\widehat{DE})$. Moreover, if $\kappa > 1$ then

$$\begin{cases} \widehat{DE}_{adj} < \widehat{DE}, & \widehat{DE} > 0, \\ \widehat{DE}_{adj} > \widehat{DE}, & \widehat{DE} < 0, \end{cases}$$

so the adjusted estimator is, in magnitude, closer to zero than the unadjusted estimator. Vice versa, if $\kappa \in (1 - \frac{1}{\bar{\pi}^e}, 1)$ then

$$\begin{cases} \widehat{DE}_{adj} > \widehat{DE}, & \widehat{DE} > 0, \\ \widehat{DE}_{adj} < \widehat{DE}, & \widehat{DE} < 0. \end{cases}$$

Thus, the adjusted estimator is, in magnitude, further away from zero than the unadjusted estimator.

A.7 Variance estimation and asymptotic distribution (no cross-fitting)

We derive variance estimators for the non-augmented (but bias-adjusted) estimators.

Indirect effect estimator. Starting with $\widehat{IE}_{adj}$. Let $r_i^a = \frac{1-p_z}{1-\pi_i^a} \left[\frac{\mathbb{I}\{\tilde{F}_i=1\}Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i=0\}Y_i}{1-p_z} \right]$ such that $\widehat{IE}_{adj} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} r_i^a$. By the ENRT design, alters in the same ego-network all have the same treatment assignments and exposure status. Thus, we can calculate the variance by using the aggregated outcomes in each ego-network. To this end, we can write

$$\begin{aligned} \widehat{IE}_{adj} &= \frac{1}{n_a} \sum_{i \in \mathcal{R}_e} \sum_{j \in \mathcal{R}_a: e(j)=i} r_j^a \\ &= \frac{1}{n_a} \sum_{i \in \mathcal{R}_e} T_i \\ &= \frac{n_e}{n_a} \bar{T}, \end{aligned}$$

where $T_i = \sum_{j \in \mathcal{R}_a: e(j)=i} r_j^a$ and $\bar{T} = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} T_i$ is the average over ego-networks. Therefore, the variance is $V_{\mathbf{Z}}(\widehat{IE}_{adj}) = \frac{n_e^2}{n_a^2} V_{\mathbf{Z}}(\bar{T})$, which can be conservatively estimated with the design-based variance estimator (Ding, 2024)

$$\widehat{V}_{\mathbf{Z}}(\widehat{IE}_{adj}) = \frac{1}{n_a^2} \sum_{i \in \mathcal{R}_e} (T_i - \bar{T})^2.$$

Under standard regularity conditions, the estimator $\widehat{IE}_{adj}$ is asymptotically normally distributed with mean IE (Ding, 2024). This motivates a Wald-type confidence interval for IE based on the variance estimator $\widehat{V}_{\mathbf{Z}}(\widehat{IE}_{adj})$

$$\widehat{IE}_{adj} \pm z_{1-\alpha/2} \sqrt{\widehat{V}_{\mathbf{Z}}(\widehat{IE}_{adj})},$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ upper quantile of a standard normal distribution. However, since the estimator $\widehat{V}_{\mathbf{Z}}(\widehat{IE}_{adj})$ is a conservative estimator of $V_{\mathbf{Z}}(\widehat{IE}_{adj})$, constructing confidence intervals based on it results in a conservative confidence interval with coverage rate of at least $1 - \alpha$.

Direct effect estimator. Moving to $\widehat{DE}_{adj}$. Let $r_i^e = \frac{\mathbb{I}\{Z_i=1\}Y_i}{p_z} - \frac{\mathbb{I}\{Z_i=0\}Y_i}{1-p_z}$ such that $\widehat{DE}_{adj} = \frac{1}{u_e} \sum_{i \in \mathcal{R}_e} r_i^e$. By the ENRT design, treatment assignment is independent across egos. We can write the variance of $\widehat{DE}_{adj}$ as

$$V_{\mathbf{Z}}(\widehat{DE}_{adj}) = \frac{1}{u_e^2} \sum_{i \in \mathcal{R}_e} V_{\mathbf{Z}}(r_i^e) + \frac{1}{u_e^2} \sum_{i,j \in \mathcal{R}_e: i \neq j} \text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e), \quad (\text{A.10})$$

where $V_{\mathbf{Z}}(r_i^e)$ is the variance of the terms r_i^e w.r.t. the design $\mathbf{Z}$. The Neyman-type variance estimator is (Ding, 2024)

$$\widehat{V}_{Neyman}(\widehat{DE}_{adj}) = \frac{1}{u_e^2} \sum_{i \in \mathcal{R}_e} (r_i^e - \widehat{DE}_{adj})^2.$$

This estimator assumes that the terms r_i^e are independent, and is therefore a conservative estimator of the first term in $V_{\mathbf{Z}}(\widehat{DE}_{adj})$. However, in the presence of contamination, this independence assumption is violated. From the exposure mapping specification (Assumption 2) and consistency (Assumption 3), we can write the terms r_i^e as

$$r_i^e = \frac{Z_i}{p_z} [F_i Y_i(1, 1) + (1 - F_i) Y_i(1, 0)] - \frac{1 - Z_i}{1 - p_z} [F_i Y_i(0, 1) + (1 - F_i) Y_i(0, 0)],$$

where $F_i = F(\mathbf{Z}_{-i}, \mathbf{A})$ are the true exposures of ego i . Therefore, if egos i and j share an ego neighbor k in the population network $\mathbf{A}$, the terms r_i^e and r_j^e will be dependent, as Z_k will influence both F_i and F_j . These dependencies induce pairwise covariances $\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e)$ that are omitted by the Neyman-type estimator $\widehat{V}_{Neyman}(\widehat{DE}_{adj})$. Thus, $\widehat{V}_{Neyman}(\widehat{DE}_{adj})$ captures only the sum of marginal variances, and ignoring the covariance terms that arise under contamination might lead to underestimation of the overall variance.

To ensure valid inference, we propose a conservative variance correction that accounts for these latent correlations. The terms r_i^e and r_j^e for two egos $i, j \in \mathcal{R}_e$ are correlated if

the egos share a common (latent) neighbor $k \in \mathcal{R}_e$. Let $K_{ij}^{\mathbf{A}}(k) = \mathbb{I}\{A_{ik} = 1, A_{jk} = 1\}$ be the indicator that in the population network $\mathbf{A}$, the ego $k \in \mathcal{R}_e$ is a common neighbor of both i and j while $k \neq i, j$. Define the number of shared neighbors between ego i and j by in the population network $\mathbf{A}$ by

$$K_{ij}^{\mathbf{A}} = \sum_{k \in \mathcal{R}_e, k \neq i, j} K_{ij}^{\mathbf{A}}(k).$$

We construct a dependency graph $\mathbf{G}_{dep}^{\mathbf{A}} = (\mathcal{R}_e, \mathbf{E}_{dep}^{\mathbf{A}})$ where an edge exists between two egos $i, j \in \mathcal{R}_e$ if the terms r_i^e and r_j^e are dependent. That is, $\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1$ occurs if $A_{ik} = A_{jk} = 1$ for some other ego $k \in \mathcal{R}_e$. Thus, we can write $\mathbb{I}\{\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1\} = \mathbb{I}\{K_{ij}^{\mathbf{A}} > 0\}$. However, the ego-ego edges in $\mathbf{A}$ are missing from the observed network $\tilde{\mathbf{A}}$. But we can compute the probability of the dependence under the sensitivity model (Section 3.4). Each of the missing ego-ego edges A_{ik} is postulated to exist independently with probability ρ_{ik}^e . Therefore, taking expectation with regard to the epistemic uncertainty from the sensitivity model yields,

$$\begin{aligned} \Pr(\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1) &= \Pr(K_{ij}^{\mathbf{A}} > 0) \\ &= 1 - \Pr(K_{ij}^{\mathbf{A}} = 0) \\ &= 1 - \prod_{k \in \mathcal{R}_e: k \neq i, j} (1 - \rho_{ik}^e \rho_{jk}^e) \\ &= \xi_{ij}. \end{aligned}$$

As the network $\mathbf{A}$ is assumed to be fixed, but partly unknown, the expressions $\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e)$ in (A.10) are essentially conditioned on it. That is, written explicitly, $\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e)$ should be viewed as $\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e \mid \mathbf{A})$. As the terms r_i^e and r_j^e are dependent only under the event $\mathbb{I}\{\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1\}$, for a fixed population network $\mathbf{A}$, we can write $\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e \mid \mathbf{A})$ as $\mathbb{I}\{\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1\} \text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e \mid \mathbf{A})$. By the Cauchy-Schwarz inequality, we can bound the covariance term

$$\mathbb{I}\{\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1\} |\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e \mid \mathbf{A})| \leq \mathbb{I}\{\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1\} \sqrt{V_{\mathbf{Z}}(r_i^e) V_{\mathbf{Z}}(r_j^e)},$$

where $V_{\mathbf{Z}}(r_i^e)$ is the variance of the terms r_i^e w.r.t. the design $\mathbf{Z}$. Since $\mathbb{I}\{\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1\}$ is unknown due to the missing ego-ego edges in the observed network, we replace it with its expected value $\Pr(\mathbf{E}_{dep}^{\mathbf{A}}(i, j) = 1) = \xi_{ij}$ under the sensitivity model. Therefore,

$$\xi_{ij} |\text{Cov}_{\mathbf{Z}}(r_i^e, r_j^e \mid \mathbf{A})| \leq \xi_{ij} \sqrt{V_{\mathbf{Z}}(r_i^e) V_{\mathbf{Z}}(r_j^e)}$$

We use the squared residuals $\hat{v}_i = (r_i^e - \widehat{DE}_{adj})^2$ as conservative estimators for the unit-level variances $V_{\mathbf{Z}}(r_i^e)$. This yields a conservative estimator for the variance in $\widehat{DE}_{adj}$ due to contamination between ego-networks

$$\widehat{V}_{Conta.}(\widehat{DE}_{adj}) = \frac{1}{u_e^2} \sum_{i \neq j} \xi_{ij} \sqrt{\hat{v}_i \hat{v}_j}. \quad (\text{A.11})$$

The final corrected variance estimator for $\widehat{DE}_{adj}$ is given by

$$\widehat{V}_{\mathbf{Z}}(\widehat{DE}_{adj}) = \widehat{V}_{Neyman}(\widehat{DE}_{adj}) + \widehat{V}_{Conta.}(\widehat{DE}_{adj}). \quad (\text{A.12})$$

We establish the asymptotic normality of $\widehat{DE}_{adj}$ by adapting standard asymptotic normality results for design-based estimators (Ding, 2024) to the dependent data setting induced by our sensitivity model. We present a sketch of a proof based on the Central Limit Theorem for dependency graphs established by Baldi and Rinott (1989). Under the design-based framework, statistical dependence between the summands r_i^e in $\widehat{DE}_{adj}$ arises solely from the random treatment assignments of shared neighbors in the latent population network. Specifically, the dependence arises from the true exposure values $F_i = F_i(\mathbf{Z}_{-i}, \mathbf{A}_i)$ due to possible contamination under the sensitivity model.

Consider the dependency graph $\mathbf{G}_{dep}^{\mathbf{A}} = (\mathcal{R}_e, \mathbf{E}_{dep}^{\mathbf{A}})$ presented above. Define the degree of each ego $i \in \mathcal{R}_e$ in this graph by $\text{Deg}_{n_e, i}$, and let $\text{Deg}_{n_e, max} = \max_{i \in \mathcal{R}_e} \text{Deg}_{n_e, i}$ denote the maximum degree. Assume that the terms r_i^e are bounded, and that the variance of the sum converges linearly n_e , i.e., $V_{\mathbf{Z}}(\sum_{i \in \mathcal{R}_e} r_i^e) = \Theta(n_e)$, indicating that the variance of the sum is bounded both above and below by n_e asymptotically. Therefore, Corollary 2 of Baldi and Rinott (1989) implies that the error in the normal approximation is bounded by a term of order $\text{Deg}_{n_e, max}/n_e^{1/4}$. Consequently, the validity of the normal approximation requires that the size of the dependency neighborhoods grows sufficiently slowly relative to the sample size, specifically $\text{Deg}_{n_e, max} = o(n_e^{1/4})$.

In the context of our sensitivity analysis, as the true population network $\mathbf{A}$ is only partially known, this imposes a constraint on the postulated latent network in the specified sensitivity model. If the specified probabilities ρ_{ij}^e are sufficiently large such that the latent network become dense (e.g., $\rho_{ij}^e \approx 1$ for all egos $i, j \in \mathcal{R}_e$), this causes the maximal degree of the dependency graph to scale linearly with the sample size ($\text{Deg}_{n_e, max} \approx n_e$), violating the condition $\text{Deg}_{n_e, max} = o(n_e^{1/4})$. Therefore, the reported confidence intervals are valid only for values of the sensitivity parameters consistent with a sufficiently sparse latent network topology.

Following a similar argument to $\widehat{IE}_{adj}$, the Wald-type confidence interval for DE based on the conservative variance estimator $\widehat{V}_{\mathbf{Z}}(\widehat{DE}_{adj})$

$$\widehat{DE}_{adj} \pm z_{1-\alpha/2} \sqrt{\widehat{V}_{\mathbf{Z}}(\widehat{DE}_{adj})},$$

results in a conservative confidence interval for DE with coverage of at least $1 - \alpha$.

A.8 Augmented estimation via cross-fitting

The augmented estimators combine the outcome Y_i and predicted outcome models for estimating direct and indirect effects. Estimating the outcome models requires cross-fitting. We utilize a two-fold cross-fitting algorithm tailored for design-based estimators (Lu et al., 2025). We also provide variance estimators for the augmented estimators obtained from this cross-fitting procedure. Finally, we describe the asymptotic distribution of the estimators.

We adapt Lu et al. (2025, Algorithm 1) to our setting. Specifically, from Assumption 1 of Bernoulli design for each ego-network, Lu et al. (2025, Proposition 1) motivates the following two-fold cross-fitting algorithm.

1. Randomly split ego-networks into two disjoint parts S_0, S_1 such that $|S_0 \cup S_1| = n$ and $S_0 \cap S_1 = \emptyset$. Specifically, we assign each ego-network (ego and alters) into S_1 with probability 0.5.
2. For $q = 0, 1$ do

- (a) Estimate the outcome functions for alters $\widehat{\mu}_i^a(f)$ and egos $\widehat{\mu}_i^e(z)$ using data from the complementary split S_{1-q} .
- (b) Predict $\widehat{\mu}_i^a(f)$ and $\widehat{\mu}_i^e(z)$ on the units from S_q .
- (c) Estimate IE on units from S_q :

$$\widehat{IE}_{adj[q]}^{aug} = \frac{1}{n_{a[q]}} \sum_{i \in \mathcal{R}_a \cap S_q} \frac{1-p_z}{1-\pi_i^a} [D_i^a + \widehat{\mu}_i^a(1) - \widehat{\mu}_i^a(0)],$$

where $n_{a[q]} = \sum_{i \in \mathcal{R}_a} \mathbb{I}\{i \in S_q\}$ is the number of alters in split S_q , and where

$$D_i^a = \frac{\mathbb{I}\{\widetilde{F}_i = 1\} (Y_i - \widehat{\mu}_i^a(1))}{p_z} - \frac{\mathbb{I}\{\widetilde{F}_i = 0\} (Y_i - \widehat{\mu}_i^a(0))}{1-p_z}.$$

- (d) Estimate DE on units from S_q :

$$\widehat{DE}_{adj[q]}^{aug} = \frac{1}{u_{e[q]}} \sum_{i \in \mathcal{R}_e \cap S_q} [D_i^e + \widehat{\mu}_i^e(1) - \widehat{\mu}_i^e(0)],$$

where $u_{e[q]} = n_{e[q]} [1 + \bar{\pi}^{e[q]}(\kappa - 1)]$ similarly to u_e in Section 3.2, with $n_{e[q]} = \sum_{i \in \mathcal{R}_e} \mathbb{I}\{i \in S_q\}$ and $\bar{\pi}^{e[q]} = \frac{1}{n_{e[q]}} \sum_{i \in \mathcal{R}_e \cap S_q} \pi_i^e$ are the number of egos and the average exposure probability, respectively, in split S_q , and

$$D_i^e = \frac{\mathbb{I}\{Z_i = 1\} (Y_i - \widehat{\mu}_i^e(1))}{p_z} - \frac{\mathbb{I}\{Z_i = 0\} (Y_i - \widehat{\mu}_i^e(0))}{1-p_z}.$$

3. Combine the estimates across splits

$$\begin{aligned} \widehat{IE}_{adj}^{aug} &= \sum_{q=0,1} \frac{n_{a[q]}}{n_a} \widehat{IE}_{adj[q]}^{aug}, \\ \widehat{DE}_{adj}^{aug} &= \sum_{q=0,1} \frac{u_{e[q]}}{u_e} \widehat{DE}_{adj[q]}^{aug}. \end{aligned}$$

Note that the weighting in the combination of the direct effect estimators uses u_e (as defined in Section 3.2) since $u_e = u_{e[0]} + u_{e[1]}$. As the units in the splits S_0 and S_1 are independent (w.r.t. $\mathbf{Z}$), variance estimation will also be a combination of variance estimates in each split (Lu et al., 2025). For the indirect effect estimates, we can write the variance estimate in split q by

$$\widehat{V}_{\mathbf{Z}} \left(\widehat{IE}_{adj[q]}^{aug} \right) = \frac{1}{n_{a[q]}^2} \sum_{i \in \mathcal{R}_e \cap S_q} \left(T_i^{aug} - \bar{T}_{[q]}^{aug} \right)^2,$$

where $T_i^{aug} = \sum_{j \in \mathcal{R}_a \cap S_q : e(j)=i} \frac{1-p_z}{1-\pi_j^a} D_j^a$ is the average of weighted residuals D_j^a of alter j in ego-network of ego $i \in \mathcal{R}_e$ in split q , and $\bar{T}_{[q]}^{aug} = \frac{1}{n_{e[q]}} \sum_{i \in \mathcal{R}_e \cap S_q} T_i^{aug}$ is the average over ego-networks in split q . The outcome model terms $\widehat{\mu}_i^e(1) - \widehat{\mu}_i^e(0)$ are not included in the variance estimator as they are fixed w.r.t. $\mathbf{Z}$ and since the splitting rule ensures independence between the two folds. The combined variance estimator of $\widehat{IE}_{adj}^{aug}$ is

$$\widehat{V}_{\mathbf{Z}} \left(\widehat{IE}_{adj}^{aug} \right) = \sum_{q=0,1} \frac{n_{a[q]}^2}{n_a^2} \widehat{V}_{\mathbf{Z}} \left(\widehat{IE}_{adj[q]}^{aug} \right). \quad (\text{A.13})$$

For direct effects, the Neyman-type variance estimator in split q is

$$\widehat{V}_{Neyman} \left(\widehat{DE}_{adj[q]}^{aug} \right) = \frac{1}{u_{e[q]}^2} \sum_{i \in \mathcal{R}_e \cap S_q} (D_i^e - \overline{D}_{[q]}^e)^2,$$

where $\overline{D}_{[q]}^e = \frac{1}{u_{e[q]}} \sum_{i \in \mathcal{R}_e \cap S_q} D_i^e$. Therefore, similarly to (A.13), the Neyman variance estimator of $\widehat{DE}_{adj}^{aug}$ is

$$\widehat{V}_{Neyman} \left(\widehat{DE}_{adj}^{aug} \right) = \sum_{q=0,1} \frac{u_{e[q]}^2}{u_e^2} \widehat{V}_{\mathbf{Z}} \left(\widehat{DE}_{adj[q]}^{aug} \right). \quad (\text{A.14})$$

We apply the same conservative contamination-attributed variance correction (A.12) derived for the adjusted estimator, adapting it to the augmented residuals. Specifically, we compute the conservative contamination-attributed variance estimator (A.11) for each fold, and combine it similarly to (A.13).

From Lu et al. (2025, Theorem 4), under common regularity conditions, both estimators $\widehat{IE}_{adj}^{aug}$ and $\widehat{DE}_{adj}^{aug}$ are approximately normal and normal-based confidence intervals using the variance estimators achieve a valid (asymptotic) coverage rate.

In addition, the random splitting minimizes the dependence between the two folds. While strict independence (w.r.t. $\mathbf{Z}$) implies no interference between the splits, the contamination between ego-networks is assumed to be sparse. Consequently, the dependence of the estimated outcome models on the treatment assignments of units in the complementary fold is expected to be negligible. Under this approximation, the estimated outcome models at each split can be treated as fixed w.r.t. the design $\mathbf{Z}$ (Lin, 2013; Lu et al., 2025). Therefore, the derivations shown in Sections A.3-A.4 imply that $\widehat{IE}_{adj}^{aug}$ and $\widehat{DE}_{adj}^{aug}$ remain approximately unbiased estimators of IE and DE , respectively, provided that the number of edges connecting units in the two folds (ego-ego and alter-ego) is small relative to the sample size.

A.9 Relative risk estimand and estimators

In the case of binary outcomes, we can also consider the relative risk (RR) estimand

$$IE^{RR} = \frac{\sum_{i \in \mathcal{R}_a} Y_i(0, 1)}{\sum_{i \in \mathcal{R}_a} Y_i(0, 0)}. \quad (\text{A.15})$$

Analogously to (3), the estimator based on the observed data is therefore

$$\widehat{IE}^{RR} = \frac{\sum_{i \in \mathcal{R}_a} \frac{\mathbb{I}\{\tilde{F}_i=1\}Y_i}{p_z}}{\sum_{i \in \mathcal{R}_a} \frac{\mathbb{I}\{\tilde{F}_i=0\}Y_i}{1-p_z}}. \quad (\text{A.16})$$

We saw in (A.1) that the numerator in (A.16) is an unbiased estimator of the numerator in (A.15). However, from (A.3) the denominator is biased. Rearranging (A.3),

$$\frac{1-p_z}{1-\pi_i^a} \mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{\tilde{F}_i=0\}Y_i}{1-p_z} \right] = Y_i(0, 0) + \frac{\pi_i^a - p_z}{1-\pi_i^a} Y_i(0, 1).$$

Thus, using (A.1), we can unbiasedly estimate $Y_i(0, 0)$ through

$$\frac{1-p_z}{1-\pi_i^a} \frac{\mathbb{I}\{\tilde{F}_i=0\}Y_i}{1-p_z} - \frac{\pi_i^a - p_z}{1-\pi_i^a} \frac{\mathbb{I}\{\tilde{F}_i=1\}Y_i}{p_z}.$$

That motivates the adjusted estimator

$$\widehat{IE}_{adj}^{RR} = \frac{\sum_{i \in \mathcal{R}_a} \frac{\mathbb{I}\{\tilde{F}_i=1\}Y_i}{p_z}}{\sum_{i \in \mathcal{R}_a} \frac{1-p_z}{1-\pi_i^a} \frac{\mathbb{I}\{\tilde{F}_i=0\}Y_i}{1-p_z} - \frac{\pi_i^a-p_z}{1-\pi_i^a} \frac{\mathbb{I}\{\tilde{F}_i=1\}Y_i}{p_z}}. \quad (\text{A.17})$$

Being a ratio estimator, (A.17) is biased, but the bias can be easily bounded (Särndal et al., 2003).

A.10 Relaxing Assumption 4

We provide two approaches to relax Assumption 4, which states that the empirical covariance between exposure probabilities π_i^e and the direct effect among exposed egos $\Delta_i(1)$ is equal to the empirical covariance between π_i^e and the direct effect among non-exposed egos $\Delta_i(0)$, that is,

$$S_{\pi^e, \Delta(1)} = S_{\pi^e, \Delta(0)}.$$

In the first approach, we derive sharp bounds for DE as a function of the sensitivity parameters π_i^e and κ and the observed data. That enables researchers to obtain bounds, instead of point estimates, for DE , or obtain bounds for DE given estimates of π_i^e and κ from validation data, as described in Section 3.6.

In the second approach, we modify the sample average proportionality expressed in (7), to a stronger assumption of unit-level proportionality. That produces a modified bias-corrected estimator for DE that is an unbiased estimator without requiring Assumption 4 to hold.

Bounds for DE without Assumption 4. We can relax Assumption 4 by using bounds instead of an unbiased estimator given values of π_i^e and κ . Note that we can write

$$S_{\pi^e, \Delta(1)} - S_{\pi^e, \Delta(0)} = S_{\pi^e, \Delta(1) - \Delta(0)}.$$

Therefore, Assumption 4 is equivalent to assuming $S_{\pi^e, \Delta(1) - \Delta(0)} = 0$. Note that from (A.9) the expectation of the naive HT estimator is

$$\begin{aligned} \mathbb{E}_Z [\widehat{DE}] &= [1 + \bar{\pi}^e(\kappa - 1)] DE + S_{\pi^e, \Delta(1)} - S_{\pi^e, \Delta(0)} \\ &= [1 + \bar{\pi}^e(\kappa - 1)] DE + S_{\pi^e, \Delta(1) - \Delta(0)} \end{aligned}$$

By the Cauchy-Schwarz inequality, we have

$$|S_{\pi^e, \Delta(1) - \Delta(0)}| \leq S_{\pi^e} S_{\Delta(1) - \Delta(0)},$$

where the first component S_{π^e} is the empirical standard deviation of π_i^e and can be computed given a specific value of the sensitivity parameters. The second standard deviation can be bounded as follows. Assume that there exist two constants $b_0 \leq b_1$ such that $b_0 \leq Y_i(z, f) \leq b_1$ for all i, z, f . Then $|\Delta_i(f)| \leq b_1 - b_0$ and $|\Delta_i(1) - \Delta_i(0)| \leq 2(b_1 - b_0)$. Therefore, by the Bhatia–Davis inequality, a sharp bound for the variance $S_{\Delta(1) - \Delta(0)}^2$ is

$$\begin{aligned} S_{\Delta(1) - \Delta(0)}^2 &\leq [2(b_1 - b_0) - (\bar{\Delta}(1) - \bar{\Delta}(0))] [\bar{\Delta}(1) - \bar{\Delta}(0) + 2(b_1 - b_0)] \\ &= 4(b_1 - b_0)^2 - (\bar{\Delta}(1) - \bar{\Delta}(0))^2. \end{aligned}$$

But from (7), we can write $\bar{\Delta}(1) - \bar{\Delta}(0) = (\kappa - 1)\bar{\Delta}(0) = (\kappa - 1)DE$. Thus, the variance bound is

$$S_{\Delta(1)-\Delta(0)}^2 \leq 4(b_1 - b_0)^2 - [(\kappa - 1)DE]^2.$$

That yields a bound for the standard deviation

$$S_{\Delta(1)-\Delta(0)} \leq \sqrt{4(b_1 - b_0)^2 - [(\kappa - 1)DE]^2} \leq 2(b_1 - b_0),$$

as $[(\kappa - 1)DE]^2 \geq 0$. Therefore, we obtain that the empirical covariance $S_{\pi^e, \Delta(1)-\Delta(0)}$ is bounded by

$$|S_{\pi^e, \Delta(1)-\Delta(0)}| \leq 2(b_1 - b_0)S_{\pi^e}.$$

Thus, we can bound DE as follows,

$$\frac{\mathbb{E}_{\mathbf{Z}} [\widehat{DE}] - 2(b_1 - b_0)S_{\pi^e}}{1 + \bar{\pi}^e(\kappa - 1)} \leq DE \leq \frac{\mathbb{E}_{\mathbf{Z}} [\widehat{DE}] + 2(b_1 - b_0)S_{\pi^e}}{1 + \bar{\pi}^e(\kappa - 1)},$$

or, since $\mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] = \frac{\mathbb{E}_{\mathbf{Z}} [\widehat{DE}]}{1 + \bar{\pi}^e(\kappa - 1)}$, we can write the bounds in terms of the bias-corrected estimator

$$\mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] - \frac{2(b_1 - b_0)S_{\pi^e}}{1 + \bar{\pi}^e(\kappa - 1)} \leq DE \leq \mathbb{E}_{\mathbf{Z}} [\widehat{DE}_{adj}] + \frac{2(b_1 - b_0)S_{\pi^e}}{1 + \bar{\pi}^e(\kappa - 1)}.$$

For example, for binary outcomes, $b_0 = 0, b_1 = 1$, and we obtain

$$\frac{\mathbb{E}_{\mathbf{Z}} [\widehat{DE}] - 2S_{\pi^e}}{1 + \bar{\pi}^e(\kappa - 1)} \leq DE \leq \frac{\mathbb{E}_{\mathbf{Z}} [\widehat{DE}] + 2S_{\pi^e}}{1 + \bar{\pi}^e(\kappa - 1)}.$$

Where in practice it is possible to use the naive estimator $\widehat{DE}$ instead of its expectation in the bounds and construct confidence intervals for the partially identified parameter DE (e.g., as proposed by Imbens and Manski, 2004).

Unit-level proportionality assumption. Recall that in the main text we defined κ as the proportion between the sample average direct effects among exposed egos $\bar{\Delta}(1)$ to the effect among non-exposed egos $\bar{\Delta}(0)$. By the design-based inference framework, this proportionality is only a notation and holds by design. The bias-corrected estimator $\widehat{DE}_{adj}$, as defined in (8), is unbiased estimator of DE only given additional conditions, as formally expressed in Proposition 3. Here, replacing Assumption 4 with a different assumption produce a modified bias-corrected estimator.

Specifically, assume that instead of constant κ that link $\Delta_i(1)$ and $\Delta_i(0)$ on average over the egos ($\bar{\Delta}(1) = \kappa\bar{\Delta}(0)$), there is a constant κ^* that control the unit-level proportionality

Assumption A.1 (Unit-level proportionality). *There exist a constant κ^* such that $\Delta_i(1) = \kappa^*\Delta_i(0)$ for all $i \in \mathcal{R}_e$.*

That is, Assumption A.1 states that $Y_i(1, 1) - Y_i(0, 1) = \kappa^* \{Y_i(1, 0) - Y_i(0, 0)\}$ for all egos. The interpretation of the additional sensitivity parameter κ in Assumption A.1 is related to the interaction between treatment and exposure in the potential outcomes of the egos. Under the saturated potential outcomes model

$$Y_i(z, f) = \beta_{0i} + \beta_{1i}z + \beta_{2i}f + \beta_{3i}zf,$$

Assumption A.1 is equivalent to $\beta_{3i} = (\kappa^* - 1)\beta_{1i}$. That is, the interaction term β_{3i} is proportional to the direct effect term β_{1i} for all egos. Thus, while κ in (7) expressed an *average-level* relation, κ^* in Assumption A.1 posit *unit-level* proportionality. However, for binary potential outcomes, Assumption A.1 implies only a specific set of possible values for κ^* , and not any constant value, thereby restricting the valid values to specify for κ^* when running the sensitivity analysis described below.

We can use a modified bias-corrected direct effect estimator

$$\widehat{DE}_{adj}^* = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \frac{1}{1 + \pi_i^e(\kappa^* - 1)} \left[\frac{\mathbb{I}\{Z_i = 1\}Y_i}{p_z} - \frac{\mathbb{I}\{Z_i = 0\}Y_i}{1 - p_z} \right].$$

Since both κ^* and π_i^e are fixed and therefore independent of $\mathbf{Z}$, we can use (A.5), (A.6), and (A.7) to obtain

$$\begin{aligned} \mathbb{E} [\widehat{DE}_{adj}^*] &= \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \frac{1}{1 + \pi_i^e(\kappa^* - 1)} [\pi_i^e \{Y_i(1, 1) - Y_i(0, 1)\} + (1 - \pi_i^e) \{Y_i(1, 0) - Y_i(0, 0)\}] \\ &\stackrel{(i)}{=} \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} \frac{1}{1 + \pi_i^e(\kappa^* - 1)} [\pi_i^e \kappa^* \{Y_i(1, 0) - Y_i(0, 0)\} + (1 - \pi_i^e) \{Y_i(1, 0) - Y_i(0, 0)\}] \\ &= \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} [Y_i(1, 0) - Y_i(0, 0)] \\ &= DE, \end{aligned}$$

where in (i) we used Assumption A.1. Therefore, $\widehat{DE}_{adj}^*$ is an unbiased estimator of DE under Assumptions 1-3 and Assumption A.1 instead of assuming homogeneous exposure probabilities or Assumption 4 in Proposition 3.

The estimation of variance for $\widehat{DE}_{adj}^*$ can be performed similarly to the estimator described in Section A.7 by replacing r_i^e with $r_i^{e*} = \frac{1}{1 + \pi_i^e(\kappa^* - 1)} \left[\frac{\mathbb{I}\{Z_i=1\}Y_i}{p_z} - \frac{\mathbb{I}\{Z_i=0\}Y_i}{1 - p_z} \right]$ such that $\widehat{DE}_{adj}^* = \frac{1}{n_e} \sum_{i \in \mathcal{R}_e} r_i^{e*}$. In addition, outcome-augmented estimation for $\widehat{DE}_{adj}^*$ is also similar to the methods described in Section A.8 for $\widehat{DE}_{adj}$.

Running the sensitivity analysis (GSA or PBA) with this modified estimator follows the same logic described in Section 3.5, by replacing κ with κ^* , albeit with a different interpretation, of course.

In addition, Assumption A.1 can be relaxed by allowing conditional proportionality given covariates $\mathbf{X}_i$. That is, replacing κ^* with $\kappa_i^* = \kappa^*(\mathbf{X}_i)$ in Assumption A.1 enables researchers to specify some structure on the proportionality instead of assuming the same level of homogeneity for all units. For example, for a categorical covariate $X_{i1} \in \{1, 2, 3\}$, researchers can specify κ_1^* for $X_{i1} = 1$ and set $\kappa_k^* = \tau_k \kappa_1^*$ for constant τ_k for $k = 2, 3$. This relaxes the homogeneous proportionality by allowing each subgroup (determined by X_{i1}) to have different values. For instance, X_{i1} can represent some proxy of risky behavior or susceptibility to be influenced by peer advice, such as age group. In that case, researchers can treat κ_1^* as the sensitivity parameters and either assume fixed values for τ_k or instead specify multiple values for τ_k as well. That can be easily done in PBA, where researchers only need to specify a distribution for these parameter values.

A.11 Estimating the sensitivity parameters with internal validation data

Alters. Assume we have an internal validation subsample of alters $\mathcal{V}_a \subset \mathcal{R}_a$ that are asked to recall information about the intervention. Assume that alters who recall such

information have a true exposure $F_i = 1$ as they were exposed to at least one treated ego. This essentially assumes that the recall data in the validation subset perfectly equate the true exposure status. This might not hold if recall is not perfect, but can be strengthened through stringent selection criteria for the validation subsample (Chao et al., 2025). The conditional probability of true exposure given observed non-exposure for alters $\Pr(F_i = 1 \mid \tilde{F}_i = 0, i \in \mathcal{R}_a)$ is linked to the marginal exposure probability π_i^a through (A.2):

$$\pi_i^a = p_z + (1 - p_z) \Pr(F_i = 1 \mid \tilde{F}_i = 0, i \in \mathcal{R}_a).$$

Assuming homogeneous exposure probabilities $\pi_i^a = \pi^a$, we can estimate this conditional probability in validation subsample through

$$\hat{q}^a = \frac{\sum_{i \in \mathcal{V}_a} \mathbb{I}\{F_i = 1, \tilde{F}_i = 0\}}{\sum_{i \in \mathcal{V}_a} \mathbb{I}\{\tilde{F}_i = 0\}},$$

and use $\hat{q}^a$ to estimate π^a

$$\hat{\pi}^a = p_z + (1 - p_z)\hat{q}^a.$$

Given $\hat{\pi}^a$, we can estimate the latent ego-alter edges probabilities ρ^a as well. By Example 1, $\pi^a = p_z + (1 - p_z) [1 - (1 - p_z \rho^a)^{n_e - 1}]$. Substituting π^a with $\hat{\pi}^a$ and solving for ρ^a yields

$$\hat{\rho}^a = \frac{1 - (1 - \hat{q}^a)^{\frac{1}{n_e - 1}}}{p_z}.$$

A similar derivation can yield an estimate of m^a , the expected number of missing ego-alter edges (Example 2).

Under heterogeneous exposure probabilities, we can regress F_i on $\tilde{F}_i = 0$ and $\mathbf{X}_i$ among the validation subsample to learn $\Pr(F_i = 1 \mid \tilde{F}_i = 0, \mathbf{X}_i, i \in \mathcal{V}_a)$ and use the learned model to predict the exposure probabilities for the other alters in the main study. In this case, it is impossible to infer the edge probabilities ρ_{ij}^a as π_i^a is a convolution of these non-identical values. However, given the estimated π_i^a values, researchers can still use the bias-adjusted indirect effect estimator (5). Accounting for uncertainty in estimating π_i^a in the final indirect effect estimates can be achieved using standard methods from the measurement error literature (e.g., as described by Chao et al., 2025).

Egos. For egos, we separate between two types of internal validation data gathered on a validation subsample $\mathcal{V}_e \subset \mathcal{R}_e$ of egos

1. Asking egos to recall *any* information related to the intervention.
2. Asking egos if they heard or learned about the intervention from other friends or sources.

The first type is similar to the validation data described for the alters. The second is more directly related to indirect exposure to treatment through neighbors.

Treated egos ($Z_i = 1$) in the validation subset $\mathcal{V}_e$ of the first type should recall any information related to the intervention, as they were randomized to participate in it. However, non-treated egos ($Z_i = 0$) would have learned about the intervention only through their relations to other treated egos. Thus, any non-treated ego that recalled information about the intervention in the validation subsample can be viewed as having been truly exposed to

at least one treated ego, that is, $F_i = 1$. By the experimental design (Assumption 1) and the exposure mapping specification (Assumption 2), the treatment status Z_i is independent of the exposure F_i . Therefore, $\pi_i^e = \Pr(F_i = 1 \mid i \in \mathcal{R}_e) = \Pr(F_i = 1 \mid Z_i = 0, i \in \mathcal{R}_e)$. Under homogeneous exposure probabilities $\pi_i^e = \pi^e$ we can estimate π_i^e from the validation subsample through

$$\hat{q}^e = \widehat{\Pr}(F_i = 1 \mid Z_i = 0, i \in \mathcal{V}_e) = \frac{\sum_{i \in \mathcal{V}_e} \mathbb{I}\{F_i = 1, Z_i = 0\}}{\sum_{i \in \mathcal{V}_e} \mathbb{I}\{Z_i = 0\}}.$$

Similarly for the derivation given to the alters, we can use $\hat{q}^e$ to estimate ρ^e (in Example 1) or m^e (in Example 2). We can also estimate heterogeneous exposure probabilities by learning $\Pr(F_i = 1 \mid Z_i = 0, \mathbf{X}_i, i \in \mathcal{V}_e)$.

In the second validation data type, we have access to the true exposure status F_i of all egos in the validation subsample. Thus, we can learn π^e or π_i^e from the entire validation subsample, instead of only non-treated egos, as described above. Moreover, κ can also be learned on this validation subsample, given that researchers are willing to assume that the validation subsample is representative of the full sample and does not suffer from selection challenges. Recall that $\kappa = \frac{\bar{\Delta}(1)}{\bar{\Delta}(0)}$. Therefore, an estimator of κ from the validation subsample is

$$\hat{\kappa} = \frac{\bar{Y}^{\mathcal{V}_e}(1, 1) - \bar{Y}^{\mathcal{V}_e}(0, 1)}{\bar{Y}^{\mathcal{V}_e}(1, 0) - \bar{Y}^{\mathcal{V}_e}(0, 0)},$$

where $\bar{Y}^{\mathcal{V}_e}(z, f) = \frac{\sum_{i \in \mathcal{V}_e} \mathbb{I}\{Z_i = z, F_i = f\} Y_i}{\sum_{i \in \mathcal{V}_e} \mathbb{I}\{Z_i = z, F_i = f\}}$, is the average outcome in the validation subsample for egos with treatment $z \in \{0, 1\}$ and exposure $f \in \{0, 1\}$. Given the estimators of π_i^e and κ from the validation subset, we can either obtain point estimates of DE using $\widehat{DE}^{adj}$ or have bounds for DE , as described in Appendix A.10.

B Extension to three-level exposure mapping

B.1 Preliminaries

We extend the framework to a three-level exposure mapping for alters, $F_i \in \{0, 1, 2+\}$, indicating exposure to zero, exactly one, or two-or-more treated neighbors, respectively. We consider only the indirect effect estimand and estimators. We derive bias expressions for the naive HT estimator and propose an adjusted estimator to correct for the bias. The adjusted estimators rely on the same sensitivity parameter of alter-ego edges probabilities ρ_{ij}^a (12) defined in Section 3.4. However, we require an additional sensitivity parameter, similar to the unit-level interaction parameter κ^* for the bias-corrected direct effect estimator presented in Appendix A.10.

Let $S_i = \sum_{j \neq i} Z_j A_{ij}$ be the true number of treated neighbors of unit i in the population network $\mathbf{A}$. For alters $i \in \mathcal{R}_a$, we can write $S_i = \sum_{j \in \mathcal{R}_e} Z_j A_{ij}$. We first modify the exposure mapping (Assumption 2).

Assumption B.1 (Three-Level Exposure Mapping). *Let*

$$F(\mathbf{Z}_{-i}, \mathbf{A}) = \begin{cases} 0, & S_i = 0 \\ 1, & S_i = 1 \\ 2+, & S_i \geq 2 \end{cases}$$

For any $\mathbf{z}, \mathbf{z}' \in \mathcal{Z}_{\mathcal{R}}$, if $z_i = z'_i$ and $F(\mathbf{z}_{-i}, \mathbf{A}) = F(\mathbf{z}'_{-i}, \mathbf{A})$, then $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$.

By Assumption B.1, we can write $Y_i(\mathbf{Z}) = Y_i(Z_i, F_i)$, where $Z_i \in \{0, 1\}$ and $F_i \in \{0, 1, 2+\}$. Thus, each unit now has six potential outcomes, instead of four. The consistency assumption (Assumption 3) is also modified.

Assumption B.2 (Consistency (Three-Level)). *The observed outcomes are given by*

$$Y_i = \sum_{z \in \{0,1\}} \sum_{f \in \{0,1,2+\}} Y_i(z, f) \mathbb{I}\{Z_i = z, F_i = f\}.$$

We focus on the estimand of the sample average indirect effect of exposure to a single versus no treated neighbors (1)

$$IE = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} Y_i(0, 1) - Y_i(0, 0).$$

B.2 Bias of the naive estimator

The naive HT estimator $\widehat{IE}$ remains as defined in (3), as the observed exposures $\tilde{F}_i$ are zero or one. Its expectation is $\mathbb{E}_{\mathbf{Z}} [\widehat{IE}] = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} (E_{i,1} - E_{i,0})$, where

$$E_{i,1} = \mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} \right]$$

$$E_{i,0} = \mathbb{E}_{\mathbf{Z}} \left[\frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right].$$

The issue is that observed exposures $\tilde{F}_i = 1$ can correspond to true exposures of either $F_i = 1$ or $F_i = 2+$, and observed exposures $\tilde{F}_i = 0$ can correspond to true exposures of $F_i = 0$, $F_i = 1$, or $F_i = 2+$. Therefore, the observed outcomes Y_i under observed exposures $\tilde{F}_i$ are mixtures of potential outcomes under different true exposures F_i .

Define $S_{i,-e(i)} = \sum_{j \in \mathcal{R}_e^{e(i)}} Z_j A_{ij}$ to be the number of treated neighbors of alter i other than its own ego $e(i)$. This represents the possible exposure from contamination between the ego-networks.

To derive the bias, we first write $E_{i,0}$ and $E_{i,1}$ as functions of the potential outcomes and the exposure probabilities, which in turn are used to obtain an expression for $\mathbb{E}_{\mathbf{Z}} [\widehat{IE}]$.

Derivation of $E_{i,1}$: If $\tilde{F}_i = 1$, then $Z_{e(i)} = 1$. The true number of treated neighbors is $S_i = 1 + S_{i,-e(i)}$. Given that $Z_{e(i)} = 1$,

- The true exposure is $F_i = 1$ if $S_i = 1$, which implies $S_{i,-e(i)} = 0$.
- The true exposure is $F_i = 2+$ if $S_i \geq 2$, which implies $S_{i,-e(i)} \geq 1$.

By Assumptions B.1-B.2,

$$\begin{aligned} E_{i,1} &= \frac{1}{p_z} \mathbb{E}_{\mathbf{Z}} \left[\mathbb{I}\{\tilde{F}_i = 1, F_i = 1\} Y_i(0, 1) + \mathbb{I}\{\tilde{F}_i = 1, F_i = 2+\} Y_i(0, 2+) \right] \\ &= \frac{1}{p_z} \mathbb{E}_{\mathbf{Z}} \left[\mathbb{I}\{Z_{e(i)} = 1, S_{i,-e(i)} = 0\} Y_i(0, 1) + \mathbb{I}\{Z_{e(i)} = 1, S_{i,-e(i)} \geq 1\} Y_i(0, 2+) \right] \\ &= \Pr(S_{i,-e(i)} = 0) Y_i(0, 1) + \Pr(S_{i,-e(i)} \geq 1) Y_i(0, 2+). \end{aligned}$$

The last step holds because $S_{i,-e(i)}$ is independent of $Z_{e(i)}$ under Assumption 1, and $\Pr(Z_{e(i)} = 1) = p_z$.

Derivation of $E_{i,0}$: If $\tilde{F}_i = 0$, then $Z_{e(i)} = 0$. The true number of treated neighbors is $S_i = S_{i,-e(i)}$. Given that $Z_{e(i)} = 0$,

- The true exposure is $F_i = 0$ if $S_i = 0$, which implies $S_{i,-e(i)} = 0$.
- The true exposure is $F_i = 1$ if $S_i = 1$, which implies $S_{i,-e(i)} = 1$.
- The true exposure is $F_i = 2+$ if $S_i \geq 2$, which implies $S_{i,-e(i)} \geq 2$.

Therefore,

$$\begin{aligned} E_{i,0} &= \frac{1}{1-p_z} \mathbb{E}_{\mathbf{Z}} \left[\sum_{f \in \{0,1,2+\}} \mathbb{I}\{\tilde{F}_i = 0, F_i = f\} Y_i(0, f) \right] \\ &= \Pr(S_{i,-e(i)} = 0) Y_i(0, 0) + \Pr(S_{i,-e(i)} = 1) Y_i(0, 1) + \Pr(S_{i,-e(i)} \geq 2) Y_i(0, 2+). \end{aligned}$$

Exposures probabilities. As in Section 3.4, the sensitivity model assumes conditional edge independence, with ρ_{ij}^a defined in (12). The random variable $S_{i,-e(i)}$ follows a Poisson-Binomial distribution with $n_e - 1$ probabilities $\{p_z \rho_{ij}^a\}_{j \in \mathcal{R}_e^{e(i)}}$. Let $\pi_{i,k}^* = \Pr(S_{i,-e(i)} = k)$. The required probabilities are

$$\begin{aligned} \pi_{i,0}^* &= \Pr(S_{i,-e(i)} = 0) = \prod_{j \in \mathcal{R}_e^{e(i)}} (1 - p_z \rho_{ij}^a) \\ \pi_{i,1}^* &= \Pr(S_{i,-e(i)} = 1) = \sum_{j \in \mathcal{R}_e^{e(i)}} \left[p_z \rho_{ij}^a \prod_{l \in \mathcal{R}_e^{e(i)} \setminus \{j\}} (1 - p_z \rho_{il}^a) \right] \\ \pi_{i,2+}^* &= \Pr(S_{i,-e(i)} \geq 2) = 1 - \pi_{i,0}^* - \pi_{i,1}^*. \end{aligned}$$

For large n_e , analytic computation of $\pi_{i,1}^*$ can be challenging, but approximation methods exist (Hong, 2013).

Bias expression. First, note that $\Pr(S_{i,-e(i)} \geq 1) = 1 - \Pr(S_{i,-e(i)} = 0) = 1 - \pi_{i,0}^*$. Substituting these probabilities into the expressions we obtained above for $E_{i,0}$ and $E_{i,1}$, the expectations of the two components of the naive estimator are:

$$\begin{aligned} E_{i,1} &= \pi_{i,0}^* Y_i(0, 1) + (1 - \pi_{i,0}^*) Y_i(0, 2+) \\ E_{i,0} &= \pi_{i,0}^* Y_i(0, 0) + \pi_{i,1}^* Y_i(0, 1) + (1 - \pi_{i,0}^* - \pi_{i,1}^*) Y_i(0, 2+) \end{aligned}$$

The expectation of the naive estimator is:

$$\begin{aligned} \mathbb{E}_{\mathbf{Z}} [\widehat{IE}] &= \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} (E_{i,1} - E_{i,0}) \\ &= \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} [-\pi_{i,0}^* Y_i(0, 0) + (\pi_{i,0}^* - \pi_{i,1}^*) Y_i(0, 1) + \pi_{i,1}^* Y_i(0, 2+)] \\ &= IE + \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} [\pi_{i,1}^* \{Y_i(0, 2+) - Y_i(0, 1)\} - (1 - \pi_{i,0}^*) \{Y_i(0, 1) - Y_i(0, 0)\}], \end{aligned} \tag{B.1}$$

which implies that the naive estimator is also biased for this indirect effect estimand. The bias depends on the average of the unit-level indirect effect of 2+ exposures relative to 1 exposures and of 1 exposures relative to 0 exposures, weighted by the exposure probabilities $\pi_{i,1}^*$ and $1 - \pi_{i,0}^*$, respectively.

B.3 Bias-corrected estimator

The resulting expected value (B.1) is a linear combination of the unknown potential outcomes $\{Y_i(0, 0), Y_i(0, 1), Y_i(0, 2+)\}$. The problem is that the observed data do not provide enough information to identify all three potential outcomes for each alter $i \in \mathcal{R}_a$. To achieve identification, we introduce an additional sensitivity parameter δ via a unit-level structural assumption on the potential outcomes, analogous to Assumption A.1 of unit-level proportionality for the egos.

Assumption B.3. *There exists a constant δ such that $Y_i(0, 2+) - Y_i(0, 1) = \delta \{Y_i(0, 1) - Y_i(0, 0)\}$ for all alters $i \in \mathcal{R}_a$.*

The sensitivity parameter δ in Assumption B.3 provides an intuitive way to model the dose-response relationship of alter exposure. It is best interpreted through the *marginal effect of additional exposure*. The effect of first exposure is $Y_i(0, 1) - Y_i(0, 0)$. The *additional* effect of the second-and-higher exposure is $Y_i(0, 2+) - Y_i(0, 1)$. Assumption B.3 implies a direct relationship. Researchers can therefore specify plausible values for δ based on their domain knowledge.

- $\delta = 0$. The effect *saturates* after one exposure, and any further contamination is irrelevant. This recovers the binary exposure model.
- $\delta > 0$. There is a *reinforcing* cumulative dose effect, where a second exposure adds benefit. For example, $\delta = 1$ implies the same marginal effect between first and second-and-higher exposure. However, $\delta \in (0, 1)$ implies that the marginal effect between first and zero exposure is higher than the marginal effect between second-or-higher and first exposure.
- $\delta < 0$. Additional exposures *diminish* the effect of the first. For example, exposure to more than one treated neighbor may create confusion or mistrust, reducing the benefit of exposure.

This framework allows researchers to conduct sensitivity analysis by specifying a plausible range for the marginal effect of exposures, rather than the absolute value of $Y_i(0, 2+)$ (e.g., by deriving sharp, yet informative, bounds). Rearranging the statement in Assumption B.3, we have

$$Y_i(0, 2+) = (\delta + 1)Y_i(0, 1) - \delta Y_i(0, 0). \quad (\text{B.2})$$

Using (B.2), the expectation in (B.1) can be written as,

$$\mathbb{E}_{\mathbf{Z}} [\widehat{IE}] = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} [\pi_{i,0}^* + \pi_{i,1}^* \delta] [Y_i(0, 1) - Y_i(0, 0)], \quad (\text{B.3})$$

which implies the bias-corrected estimator of indirect effect under the three-level exposures

$$\widehat{IE}_{adj} = \frac{1}{n_a} \sum_{i \in \mathcal{R}_a} \frac{1}{\pi_{i,0}^* + \pi_{i,1}^* \delta} \left[\frac{\mathbb{I}\{\tilde{F}_i = 1\} Y_i}{p_z} - \frac{\mathbb{I}\{\tilde{F}_i = 0\} Y_i}{1 - p_z} \right]. \quad (\text{B.4})$$

Following the same strategy as in the proof of Proposition 2, and utilizing the derivations above, it can be shown that the adjusted estimator $\widehat{IE}_{adj}$ is unbiased for the indirect effect estimand IE

$$\mathbb{E}_{\mathbf{Z}} [\widehat{IE}_{adj}] = IE.$$

Moreover, (B.3) implies that the naive estimator $\widehat{IE}$ is unbiased if

- $\pi_{i,0}^* = 1$. This corresponds to no contamination between ego-networks (no additional treated neighbors beyond the ego). That is equivalent to $\rho_{ij}^a = 0$ for all alter-ego edges.
- For all $i \in \mathcal{R}_a$, we have that $\delta = \frac{1-\pi_{i,0}^*}{\pi_{i,1}^*} = 1 + \frac{\pi_{i,2+}^*}{\pi_{i,1}^*}$. This implies that the marginal effect of additional exposure is exactly offset by the contamination between ego-networks.

B.4 Relation to the two-level exposure model

Note that if we set $\delta = 0$ in (B.2), then $Y_i(0, 2+) = Y_i(0, 1)$, and the additional exposure level 2+ collapses into level 1. Recall that $\pi_i^a = \Pr(F_i = 1 \mid i \in \mathcal{R}_a)$ is the probability that alter i is exposed to *at least* one treated neighbor under the two-level exposure mapping. We can write

$$\begin{aligned} 1 - \pi_i^a &= \Pr(F_i = 0 \mid Z_{e(i)} = 1 \in \mathcal{R}_a) \Pr(Z_{e(i)} = 1 \mid i \in \mathcal{R}_a) \\ &\quad + \Pr(F_i = 0 \mid Z_{e(i)} = 0 \in \mathcal{R}_a) \Pr(Z_{e(i)} = 0 \mid i \in \mathcal{R}_a) \\ &= \Pr(S_{i,-e(i)} = 0)(1 - p_z) \\ &= \pi_{i,0}^*(1 - p_z). \end{aligned}$$

That is, $\pi_{i,0}^* = \frac{1-\pi_i^a}{1-p_z}$. Substituting this into (B.4) with $\delta = 0$, we recover the bias-corrected estimator for the two-level exposure mapping given in (5). Therefore, the two-level exposure model can be viewed as a special case of the three-level exposure model with saturation of effect after the first exposure.

B.5 Direction of bias, sign preservation, and effect magnitude

If $Y_i(0, 1) - Y_i(0, 0) \geq 0$ for all alters $i \in \mathcal{R}_a$, then

$$\begin{cases} \mathbb{E}_{\mathbf{Z}} [\widehat{IE}] \geq IE \geq 0, & \delta \geq \max_{i \in \mathcal{R}_a} \left\{ 1 + \frac{\pi_{i,2+}^*}{\pi_{i,1}^*} \right\}, \\ 0 \leq \mathbb{E}_{\mathbf{Z}} [\widehat{IE}] \leq IE, & \delta \in \left(\max_{i \in \mathcal{R}_a} \left\{ 1 - \frac{1-\pi_{i,2+}^*}{\pi_{i,1}^*} \right\}, \min_{i \in \mathcal{R}_a} \left\{ 1 + \frac{\pi_{i,2+}^*}{\pi_{i,1}^*} \right\} \right), \\ \mathbb{E}_{\mathbf{Z}} [\widehat{IE}] \leq 0 \leq IE, & \delta \leq \min_{i \in \mathcal{R}_a} \left\{ 1 - \frac{1-\pi_{i,2+}^*}{\pi_{i,1}^*} \right\}. \end{cases}$$

Similarly, if $Y_i(0, 1) - Y_i(0, 0) \leq 0$ for all alters $i \in \mathcal{R}_a$, we obtain the opposite inequalities. Therefore, given monotonicity (in either direction) of the potential outcomes $Y_i(0, 1)$ and $Y_i(0, 0)$, the estimator $\widehat{IE}$ is biased away, towards, or changes sign relative to the true indirect effect IE , depending on the value of the sensitivity parameter δ .

Similarly to the augmented bias-corrected estimator (Section 3.3), we can define an augmented version of the three-level exposure bias-corrected estimator by adding outcome models $\widehat{\mu}_i^a(f)$ for $f \in \{0, 1\}$. Variance estimation can be achieved using the same approaches described in Section A.7 for the bias-corrected indirect effect estimators under two-level exposures.

B.6 Sensitivity analysis

The bias-corrected estimator (B.4) can be used in the same way as the bias-corrected estimators under two-level exposure mapping described in Section 3.5. That is, based on their domain knowledge, researchers can specify plausible ranges for the sensitivity

parameters δ and the basic components that result in ρ_{ij}^a and conduct GSA or PBA. The main difference is the specification of the additional sensitivity parameter δ , and a slightly more complicated computation of the exposure probabilities $\pi_{i,0}^*$ and $\pi_{i,1}^*$. The latter can be easily computed in **R** using the approximations developed by Hong (2013), for example, as implemented in the `poisbinom` package.

C Probabilistic bias analysis

We provide a brief overview of the technical motivation behind PBA. For a more detailed treatment, we refer the reader to Greenland (2005) and Fox et al. (2021).

Assume researchers observe data $\mathbf{D}$ and are interested in estimating a parameter θ (e.g., a causal effect). They construct an estimator $\hat{\theta}(\mathbf{D})$ based on the observed data, but know that it is biased. Namely,

$$\mathbb{E} \left[\hat{\theta}(\mathbf{D}) \right] \neq \theta.$$

Suppose there exists a vector of sensitivity or bias parameters ρ such that the researchers can construct an adjusted estimator $\hat{\theta}_{adj}(\mathbf{D}, \rho)$ that is an unbiased estimator of θ given the true value of ρ :

$$\mathbb{E} \left[\hat{\theta}_{adj}(\mathbf{D}, \rho) \mid \rho \right] = \theta.$$

However, the true values of ρ are typically unknown and not identifiable from the data. Standard GSA compares estimates of $\hat{\theta}_{adj}(\mathbf{D}, \rho)$ (and possibly associated uncertainty quantification) across a fixed grid of ρ values, treating each point in the grid as if it were the truth. In contrast, PBA treats ρ as a random variable governed by a probability distribution $p(\rho)$. This distribution reflects the researcher’s prior knowledge or uncertainty regarding the true values and structure of the bias.

The goal of PBA is to estimate the “posterior” distribution of the parameter of interest θ , accounting for both the uncertainty in ρ and the sampling variability of the estimator. This distribution can be approximated via Monte Carlo simulation (Greenland, 2005). First, we draw M samples $\rho^{(1)}, \dots, \rho^{(M)}$ from the prior $p(\rho)$. For each draw $\rho^{(m)}$, we calculate the adjusted point estimate $\hat{\theta}_{adj}(\mathbf{D}, \rho^{(m)})$ and its associated variance estimate, denoted $\hat{V}(\mathbf{D}, \rho^{(m)})$. To fully propagate the statistical uncertainty, we simulate a value $\theta^{(m)}$ from the estimator’s approximate sampling distribution, typically assumed to be normal:

$$\theta^{(m)} \sim N \left(\hat{\theta}_{adj}(\mathbf{D}, \rho^{(m)}), \hat{V}(\mathbf{D}, \rho^{(m)}) \right).$$

The resulting values $\theta^{(1)}, \dots, \theta^{(M)}$ approximate the marginal “posterior” distribution of the parameter of interest θ , accounting for both the uncertainty regarding the sensitivity parameters and the statistical uncertainty arising from the data. This distribution can be summarized through standard summary statistics (e.g., the mean and percentiles).

In our context, the only randomness in the data is due to the randomization $\mathbf{Z}$. Thus, $\hat{\theta}$ corresponds to the naive estimators $\widehat{IE}$ or $\widehat{DE}$. The sensitivity ρ consists of the edge probabilities ρ_{ij}^a or ρ_{ij}^e and the interaction parameter κ . The adjusted estimators $\hat{\theta}_{adj}$ correspond to $\widehat{IE}_{adj}$ or $\widehat{DE}_{adj}$. This formulation yields the PBA procedure described in Section 3.5.

D Simulations

Code for replicating the simulation study and data analysis is available at https://github.com/barwein/ENRT_SA.

We generated three covariates for each unit.

$$\begin{aligned} X_{i1} &\sim \text{Bernoulli}(p_1), \\ X_{i2} &\sim \text{Bernoulli}(p_2), \\ X_{i3} &\sim \text{Normal}(0, 1), \end{aligned}$$

where $p_1 = 0.6, p_2 = 0.2$ for egos and $p_1 = 0.5, p_2 = 0.3$ for alters. The covariates coefficients in the potential outcome models were set to $\beta_e = (-0.5, -0.3, 0.2)$ for egos and $\beta_a = (-0.4, -0.2, 0.1)$ for alters. Heterogeneous edge probabilities were generated with weights computed using the Euclidean distance $\|\cdot\|_2$ between the covariates of units.

Tables D.1 and D.2 report the simulation results for the indirect and direct effects, respectively, with homogeneous exposure probabilities and augmented estimators. All results are qualitatively similar to those presented in the main text.

Table D.3 shows the results of the direct effect estimators under heterogeneous exposure probabilities (with $m_e = 150$) under misspecification of κ values. That is, the true value is $\kappa = 1.5$ while we tested how using $\kappa = 0.5$ and $\kappa = 1.1$ impacts the estimates. The table shows that using $\kappa = 0.5$ severely affects the bias-corrected estimators $\widehat{DE}_{adj}$ and $\widehat{DE}_{adj}^{aug}$ by increasing the bias and reducing the coverage compared to the naive HT estimator $\widehat{DE}$. That is the case since using $\kappa < 1$ when the true value is larger than one, changes the direction of the bias correction in the weights used in (8), resulting in inferior performance to the naive estimator. However, using a misspecified value closer to the true value $\kappa = 1.1$ (which is still in the same direction $\kappa > 1$ as the true value), negatively impacts the bias and coverage of the bias-corrected estimators, but still results in better performance than the naive estimator.

Table D.4 shows the results of the direct effect estimators under heterogeneous exposure probabilities (with $m_e = 150$) with violation of Assumption 4. The violation was simulated by modifying the data-generating process of the potential outcomes. Specifically, we have added a heterogeneous interaction term such that the potential outcomes for egos $i \in \mathcal{R}_e$ are now generated from

$$Y_i(z, f) = -0.5 + 2z + 0.5f + zf + \beta_4 zf X_{i1} + \mathbf{X}_i^T \beta_e + \varepsilon_i.$$

Taking $\beta_4 = 0$ yields the original data-generating process used in the other setups. However, taking $\beta_4 \neq 0$ results in heterogeneous treatment and exposure interactions, leading to violation of Assumption 4. Table D.4 shows the results with $\beta_4 = 0.5$ and $\beta_4 = 1.1$, while using the true κ value in the bias-corrected estimators. Larger β_4 leads to stronger violation of Assumption 4. The table shows that the violation of Assumption 4 only slightly affects the bias-corrected estimators regardless of their specification and β_4 values. Specifically, the bias was minimal, but the augmented estimators tend to underestimate the variance ($SD/SE > 1$).

Finally, Tables D.5 and D.6 show the results for the indirect and direct effects, respectively, with binary potential outcomes under homogeneous or heterogeneous exposure probabilities with $m_a = 100$ and $m_e = 150$. The binary potential outcomes were sampled from a logistic model with

$$\begin{aligned} \text{logit Pr}(Y_i(z, f) = 1) &= -0.5 + 2z + 0.5f + zf + \mathbf{X}_i^T \beta_e, & \text{for } i \in \mathcal{R}_e, \\ \text{logit Pr}(Y_i(0, f) = 1) &= -0.5 + 2f + \mathbf{X}_i^T \beta_a, & \text{for } i \in \mathcal{R}_a, \end{aligned}$$

using the same values for β_e and β_a as described above. We targeted the DE and IE on the difference scale, but the outcome-augmented estimators used the predicted probabilities from a logistic regression model. The results are qualitatively similar to those with continuous potential outcomes, showing that the bias-corrected estimators performed better than the naive estimators.

Table D.1: Simulation results for indirect effect with homogeneous exposure probabilities. True effect is $IE = 2$.

Scenario	Specification	Augmented	Bias	Coverage	SD/SE
$m^a = 100$	Heterogeneous	FALSE	0.000	0.963	0.941
		TRUE	-0.001	0.954	0.986
	Homogeneous	FALSE	-0.001	0.962	0.942
		TRUE	-0.002	0.954	0.986
	Naive	FALSE	-0.236	0.506	0.942
		TRUE	-0.237	0.400	0.986
$m^a = 200$	Heterogeneous	FALSE	0.008	0.972	0.881
		TRUE	0.006	0.954	0.962
	Homogeneous	FALSE	0.003	0.973	0.882
		TRUE	0.001	0.956	0.959
	Naive	FALSE	-0.440	0.063	0.882
		TRUE	-0.442	0.020	0.959

Table D.2: Simulation results for direct effect with homogeneous exposure probabilities. True effect is $DE = 2$.

Scenario	Specification	Augmented	Bias	Coverage	SD/SE
$m^e = 150, \kappa = 1.5$	Heterogeneous	FALSE	0.012	0.984	0.798
		TRUE	0.008	0.956	0.961
	Homogeneous	FALSE	0.000	0.982	0.810
		TRUE	-0.003	0.955	0.969
	Naive	FALSE	0.529	0.247	1.079
		TRUE	0.525	0.132	1.052
$m^e = 250, \kappa = 1.5$	Heterogeneous	FALSE	0.026	0.999	0.583
		TRUE	0.023	0.980	0.827
	Homogeneous	FALSE	0.007	0.999	0.600
		TRUE	0.004	0.978	0.843
	Naive	FALSE	0.724	0.076	1.070
		TRUE	0.720	0.012	1.031

Table D.3: Simulation results for direct effect with heterogeneous exposure probabilities and misspecified κ values. True effect is $DE = 2$, true $\kappa = 1.5$, and $m_e = 150$.

Scenario	Specification	Augmented	Bias	Coverage	SD/SE
Using $\kappa = 0.5$	Heterogeneous	FALSE	1.385	0.008	0.755
		TRUE	1.376	0.000	0.951
	Homogeneous	FALSE	1.417	0.006	0.772
		TRUE	1.408	0.000	0.959
	Naive	FALSE	0.514	0.279	1.068
		TRUE	0.508	0.175	1.045
Using $\kappa = 1.1$	Heterogeneous	FALSE	0.392	0.716	0.791
		TRUE	0.385	0.430	0.952
	Homogeneous	FALSE	0.388	0.697	0.813
		TRUE	0.382	0.429	0.960
	Naive	FALSE	0.514	0.279	1.068
		TRUE	0.508	0.175	1.045

Table D.4: Simulation results for direct effect with heterogeneous exposure probabilities and violation of Assumption 4. True effect is $DE = 2$ and $m_e = 150$. The value β_4 controls the magnitude of violation, where larger values correspond to larger violations. Values of κ depend on β_4 and are computed from the potential outcomes.

Scenario	Specification	Augmented	Bias	Coverage	SD/SE
$\beta_4 = 0.5$ and $\kappa = 1.68$	Heterogeneous	FALSE	0.003	0.984	0.792
		TRUE	-0.002	0.955	0.969
	Homogeneous	FALSE	-0.010	0.981	0.815
		TRUE	-0.016	0.952	0.977
	Naive	FALSE	0.681	0.122	1.094
		TRUE	0.673	0.049	1.063
$\beta_4 = 1.5$ and $\kappa = 1.97$	Heterogeneous	FALSE	0.009	0.980	0.814
		TRUE	0.003	0.950	0.993
	Homogeneous	FALSE	-0.010	0.976	0.840
		TRUE	-0.015	0.949	1.000
	Naive	FALSE	1.013	0.020	1.134
		TRUE	1.004	0.005	1.087

Table D.5: Simulation results for indirect effect with binary potential outcomes under homogeneous or heterogeneous exposure probabilities with $m_a = 100$. Augmented estimators used a logistic regression model for the augmentation. True effect is $IE = 0.4775$.

Scenario	Specification	Augmented	Bias	Coverage	SD/SE
Heterogeneous	Heterogeneous	FALSE	0.001	0.968	0.913
		TRUE	0.001	0.984	0.789
	Homogeneous	FALSE	0.001	0.967	0.913
		TRUE	0.001	0.985	0.788
	Naive	FALSE	-0.055	0.928	0.913
		TRUE	-0.055	0.834	0.788
Homogeneous	Heterogeneous	FALSE	-0.003	0.968	0.912
		TRUE	-0.003	0.987	0.786
	Homogeneous	FALSE	-0.003	0.967	0.912
		TRUE	-0.003	0.986	0.785
	Naive	FALSE	-0.059	0.920	0.912
		TRUE	-0.059	0.804	0.785

Table D.6: Simulation results for direct effect with binary potential outcomes under homogeneous or heterogeneous exposure probabilities with $m_e = 150$. Augmented estimators used a logistic regression model for the augmentation. True effect is $DE = 0.52$ and $\kappa = 0.932$.

Scenario	Specification	Augmented	Bias	Coverage	SD/SE
Heterogeneous	Heterogeneous	FALSE	-0.005	0.994	0.674
		TRUE	-0.006	0.982	0.812
	Homogeneous	FALSE	-0.004	0.993	0.692
		TRUE	-0.005	0.981	0.818
	Naive	FALSE	-0.023	0.954	0.939
		TRUE	-0.024	0.963	0.886
Homogeneous	Heterogeneous	FALSE	-0.012	0.996	0.675
		TRUE	-0.012	0.987	0.772
	Homogeneous	FALSE	-0.011	0.994	0.689
		TRUE	-0.012	0.986	0.778
	Naive	FALSE	-0.029	0.953	0.934
		TRUE	-0.030	0.964	0.845

E Data analysis

Figure E.1 shows the distribution of the number of alters per ego-network in the data. The average and median number is 2, the range is from 1 to 6.

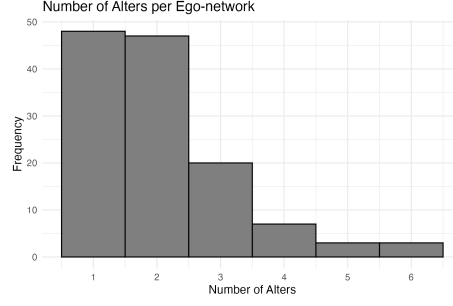


Figure E.1: Distribution of number of alters per ego-network.

Figure E.2 complements Figure 3 and shows the the GSA results (Section 5.1) for the direct effect using the augmented estimators given that $\kappa = 1.5$. We also included the results using $\gamma^e = 2$, representing a stronger influence of covariates' similarity on the exposure probabilities π_i^e . This presentation, similar to that of the indirect effect, enables us to see how changing the expected number of missing ego-ego edges (m^e) impacts the estimated direct effect. Clearly, for all specifications, the estimated effects were almost identical and approached zero as the contamination level increased. Furthermore, Figure E.3 shows the results of GSA with the non-augmented estimators $\widehat{DE}_{adj}$, also including $\gamma^e = 2$. The figure shows that under all three specifications, the bias-corrected estimates are similar. Figure E.4 complements Figure E.2 by showing the bias-corrected estimates with the non-augmented estimators for $\kappa = 1.5$. This figure highlights the similarity of the results across the different edge probability specifications. Finally, Figure E.5 shows the augmented bias-corrected DE estimates as a function of κ given the value $\widehat{m}^e = 33.4$ estimated from the internal validation data, showing that increasing κ yields bias-corrected estimates closer to zero.

Figure E.6 shows the results of the GSA for indirect effect using the non-augmented estimators $\widehat{IE}_{adj}$. The structure is similar to that of the non-augmented estimator presented in the main text (Figure 2). However, the augmented estimator had values closer to zero for all levels of contamination (including no contamination) and edge probability specifications. Taking $\gamma^e = 2$ results in higher variance estimates for large levels of missing alter-ego edges ($m^a \geq 250$).

Figures E.7 and E.8 show the results of the PBA for indirect and direct effect using the non-augmented estimators $\widehat{IE}_{adj}$ and $\widehat{DE}_{adj}$. The results are similar to those of the non-augmented estimators presented in the main text (Figure 4).

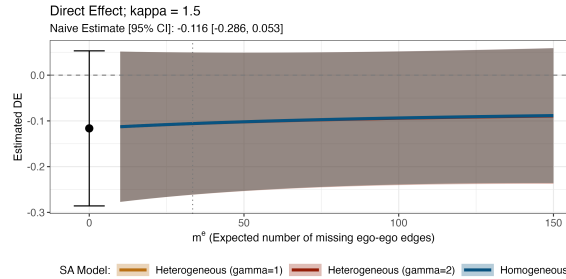


Figure E.2: GSA for direct effect with $\kappa = 1.5$ using the augmented estimators $\widehat{DE}_{adj}^{aug}$. The dashed vertical line at 33.4 is the estimate of m^e from the internal validation data.

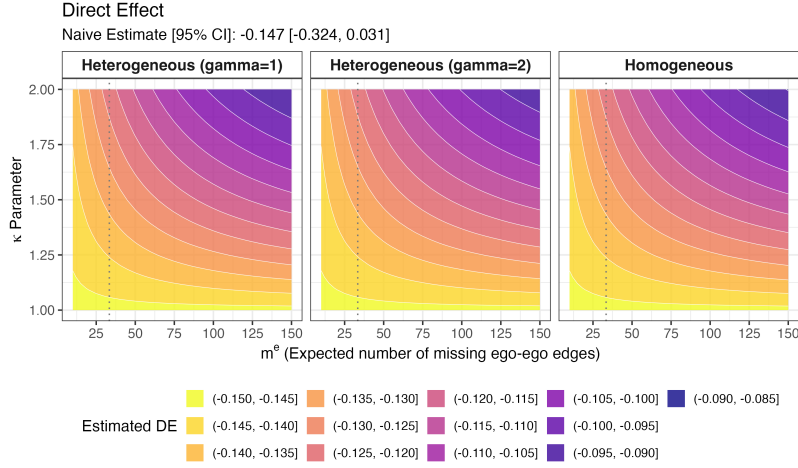


Figure E.3: GSA for direct effect with non-augmented estimators $\widehat{DE}_{adj}$. The dashed vertical line at 33.4 is the estimate of m^e from the internal validation data.

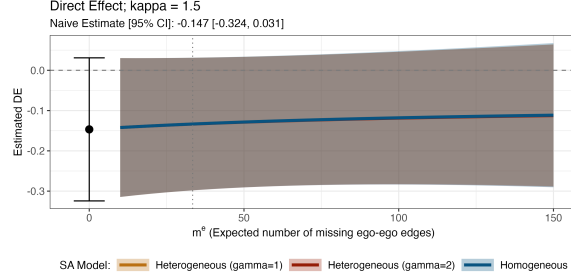


Figure E.4: GSA for direct effect with $\kappa = 1.5$. Not-augmented estimator. The dashed vertical line at 33.4 is the estimate of m^e from the internal validation data.

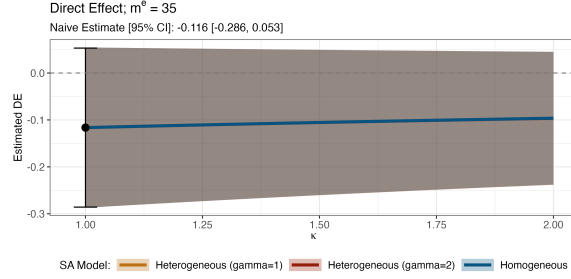


Figure E.5: GSA for direct effect with augmented estimators using $m^e = 35$ to approximate the bias-corrected estimates obtained using the the estimated value $\hat{m}^e = 33.4$ from the internal validation data.

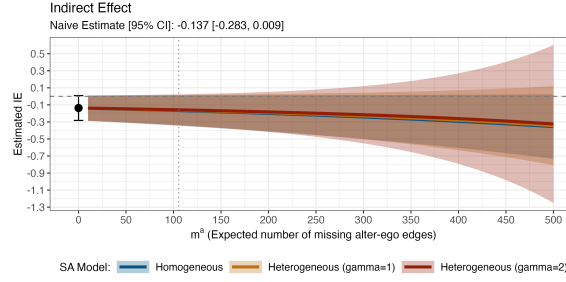


Figure E.6: GSA for indirect effect with non-augmented estimators $\widehat{IE}_{adj}$. The dashed vertical line at 105.5 is the estimate of m^a from the internal validation data.

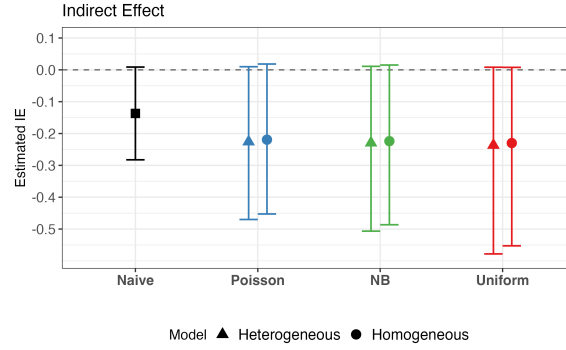


Figure E.7: PBA for indirect effect with non-augmented estimators $\widehat{IE}_{adj}$. Results are shown as mean (95% intervals) of bias-corrected estimates across $B = 10^4$ Monte Carlo samples. The results account for both bias and statistical uncertainty.

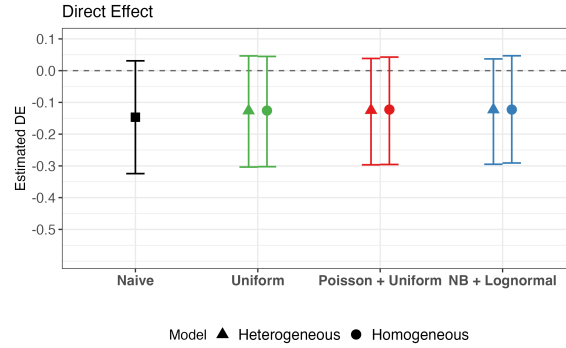


Figure E.8: Probabilistic bias analysis for direct effect with non-augmented estimators $\widehat{DE}_{adj}$. Results are shown as mean (95% intervals) of bias-corrected estimates across $B = 10^4$ Monte Carlo samples. The results account for both bias and statistical uncertainty.

F Results of the example analysis code

Analysis code applied to a simulated dataset similar to the design of HPTN 037 can be found at https://github.com/barwein/ENRT_SA. To replicate it, run the script `ENRT_example_code.R`. We provide here the results from running the code, in the order of their appearance in the script.

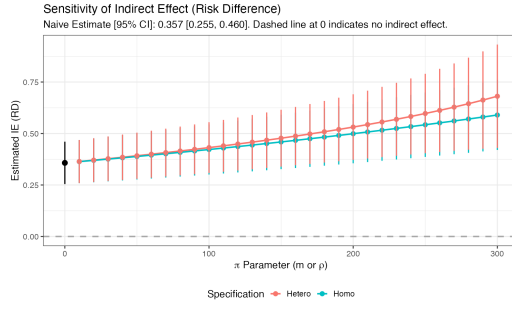


Figure F.1: Results for IE estimand with GSA in the ENRT example code.

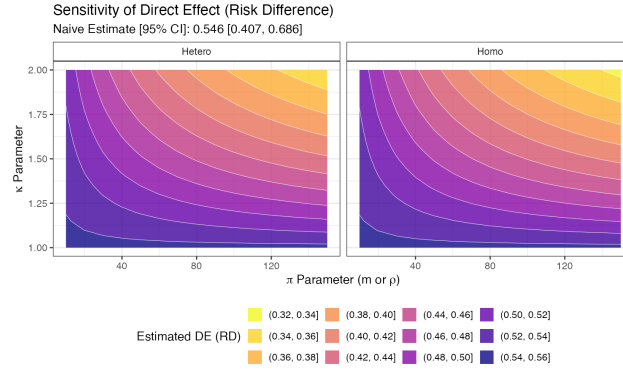


Figure F.2: Results for DE estimand with GSA in the ENRT example code.

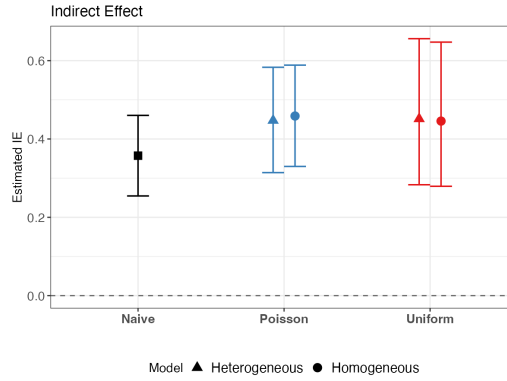


Figure F.3: Results for IE estimand with PBA in the ENRT example code.

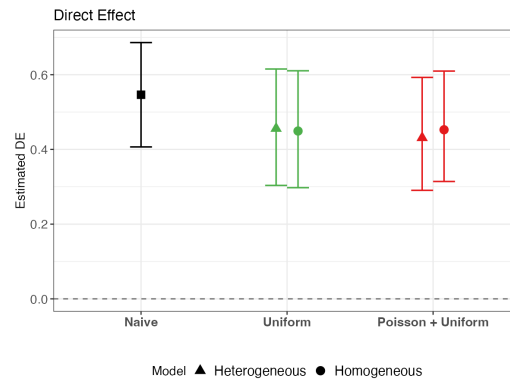


Figure F.4: Results for DE estimand with PBA in the ENRT example code.