HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2023 May 26.

Published in final edited form as:

Biometrics. 2023 June ; 79(2): 1409–1419. doi:10.1111/biom.13606.

Assessing Intervention Effects in a Randomized Trial within a Social Network

Shaina J. Alexandria[1,*], **Michael G. Hudgens**[2,**], **Allison E. Aiello**[3]

[1]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, U.S.A.

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

[3]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

SUMMARY:

Studies of social networks provide unique opportunities to assess the causal effects of interventions that may impact more of the population than just those intervened on directly. Such effects are sometimes called peer or spillover effects, and may exist in the presence of interference, i.e., when one individual's treatment affects another individual's outcome. Randomization-based inference (RI) methods provide a theoretical basis for causal inference in randomized studies, even in the presence of interference. In this article, we consider RI of the intervention effect in the eX-FLU trial, a randomized study designed to assess the effect of a social distancing intervention on influenza-like-illness transmission in a connected network of college students. The approach considered enables inference about the effect of the social distancing intervention on the per-contact probability of influenza-like-illness transmission in the observed network. The methods allow for interference between connected individuals and for heterogeneous treatment effects. The proposed methods are evaluated empirically via simulation studies, and then applied to data from the eX-FLU trial.

Keywords

Causal inference; eX-FLU; Interference; Social distancing; Spillover effect; Stochastic potential outcomes

1. Introduction

The novel coronavirus disease 2019 (COVID-19) can be spread through person-to-person contacts (Paules et al., 2020). Social distancing after symptom onset has been shown to be effective in preventing transmission of similar human coronavirus strains, including severe

* shaina.alexandria@northwestern.edu . ** mhudgens@bios.unc.edu .

SUPPORTING INFORMATION

Web Tables and Figures referenced in Section 3 are available with this paper at the Biometrics website on Wiley Online Library. Code for the simulation study can be found at github.com/shaina-alexandria/netrix.

acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) (Peak et al., 2017). Therefore, public health efforts to control the spread of COVID-19 have included isolation of individuals who exhibit symptoms (Nussbaumer-Streit et al., 2020). Given that individuals may be infectious before symptom onset and that isolation of one individual can lower the chances of infection for everyone with whom that individual would have otherwise had contact, how can the efficacy of social distancing interventions be measured?

This question is not unique to the study of coronavirus. Public health interventions are often designed to disseminate benefits to more of the population than is treated directly (Clemens et al., 2011; Klugman, 2014). This dissemination effect, called the spillover or peer effect, is an important component of the population-level effect of an intervention (Clemens et al., 2011; Halloran and Hudgens, 2012). Randomized trials conducted within social networks can provide information about the spillover effect by randomly administering the intervention to individuals within the network (Eckles et al., 2017). However, network-based trials require careful consideration when drawing inference about the effects of the intervention.

As a motivating example, consider the eX-FLU trial (Aiello et al., 2016), which followed a sample of college students during influenza season to evaluate the efficacy of a social distancing intervention for the prevention of influenza-like-illness (ILI) transmission. Enrollment in the eX-FLU trial relied on respondent-driven sampling (RDS) to ensure the study sample contained connected networks of individuals. A respondent-driven sample is formed by a chain-referral procedure in which seed individuals are selected from the target population and then provided with tokens to nominate other members of the target population (Goel et al., 2010). This process continues until the desired sample size is reached.

In general, RDS samples will be non-representative of the target population and will exhibit dependence among individuals in the sample (Goel et al., 2010). This dependence may introduce interference. Interference occurs if the intervention status of one person affects another person’s outcome (Cox, 1958; Rosenbaum, 2007). Non-random sampling, dependence, and interference all pose inferential challenges. For instance, large-sample frequentist methods typically assume random sampling from the target population (i.e., observations are independent and identically distributed), and therefore would not generally produce valid inference in settings such as the eX-FLU trial.

Many of the existing methods for causal inference in the presence of interference assume either that interference is restricted to disjoint subgroups (Sobel, 2006; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2012) or that the network is a random draw from the super-population (Liu et al., 2016; Ogburn et al., 2017; Ogburn and Vanderweele, 2017; Forastiere et al., 2021; Tchetgen Tchetgen et al., 2021). In settings such as the eX-FLU trial where such assumptions are dubious, randomization-based inference (RI) provides a natural approach to assessing intervention effects (Aronow and Samii, 2017; Athey and Imbens, 2017; Basse et al., 2019; Wu and Ding, 2020). Such an approach is considered in this paper for analysis of the eX-FLU data. The proposed approach makes

no asymptotic assumptions and allows for inference about the effect of intervention on ILI transmission in the observed network.

The outline of the remainder of this paper is as follows. Section 2 provides additional details about the eX-FLU trial and describes RI methods for inference about the effect of a social distancing intervention. The proposed approach allows for interference between connected individuals and for heterogeneous treatment effects. Section 3 presents simulation studies evaluating the methods under a variety of scenarios. Data from the eX-FLU trial is analyzed in Section 4, and some concluding remarks are given in Section 5.

2. Methods

2.1 Introduction to the eX-FLU trial and notation

The eX-FLU trial was designed to study the efficacy of a three-day social distancing intervention for prevention of ILI transmission in a network of students at a large university. Social distancing (isolation) aims to reduce contact between individuals in an effort to limit opportunities for disease transmission. The eX-FLU experiment followed the students over a ten-week period during an influenza season to track ILI transmission (Aiello et al., 2016). Of the 5536 eligible students, 590 of those identified through the RDS process elected to participate in the trial. Upon meeting the case definition for ILI, a student in the intervention group was encouraged to self-isolate in their dorm room for three days while receiving food and services from the eX-FLU staff. ILI symptoms were self-reported and confirmed with laboratory testing for influenza and other respiratory illnesses (Aiello et al., 2016). The eX-FLU study design is detailed below and notation is defined generally for any study of similar design.

Consider an experiment such as eX-FLU, conducted on n individuals who formed a network of face-to-face social contacts. In the following, random variables are denoted by uppercase letters and fixed quantities are denoted by lowercase letters. Suppose individuals were randomly assigned to intervention or control at baseline (i.e., the beginning of the trial). Assume the randomization assignment did not depend on the social network structure, risk of developing ILI, or any other individual characteristics. For $i \in \{1, \ldots, n\}$, let A_i equal 1 if individual i was assigned to the intervention group and 0 if individual i was assigned to the control condition. Let $\mathbf{A} = (A_1, \ldots, A_n)$ represent the vector of observed intervention assignments. Let $\mathcal{A}$ denote the set of all possible intervention assignment vectors $\mathbf{a} = (a_1, \ldots, a_n)$ given the randomization design. For example, eX-FLU employed a cluster randomized design, in which clusters (or groups) of students were defined by subdivisions of student residence halls, and each cluster of individuals was randomly assigned to the intervention or control condition. By design, exactly m of the h clusters were assigned to the intervention, such that $\mathcal{A}$ is composed of the $\binom{h}{m}$ possible assignment vectors where $a_i = a_j$ for all individuals i and j in the same cluster. For simplicity, assume each assignment in $\mathcal{A}$ occurs with equal probability, although this assumption can easily be relaxed. For eX-FLU, $P(\mathbf{A} = \mathbf{a}) = \binom{h}{m}^{-1}$ for all $\mathbf{a} \in \mathcal{A}$.

Suppose the study participants were then followed for τ discrete time points. During followup, information was collected on a binary outcome of interest, e.g., ILI status. Let Y_{it} denote the observed outcome for individual i at time $t \in \{1, \ldots, \tau\}$. Further, let $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{nt})$ be the observed outcomes for all n individuals at time t. In some studies, the outcome may be collected at baseline. Denote the baseline outcome for individual i by y_{i0}. In eX-FLU, 25 individuals had ILI at baseline (13 cases in the intervention group).

The eX-FLU intervention was not intended to prevent ILI in the person assigned to the intervention; instead, this study investigated the spillover effect of social distancing on at-risk peers. Therefore, information on social contacts between study participants was also collected at each of the τ time points. Self-reported contact information was collected for each individual via weekly questionnaires, and additional contact information was collected through residence hall and course rosters. In the analysis presented below, individuals i and j are considered contacts at time t, denoted by binary indicator $e_{ijt} = 1$, if either individuals i or j (or both) had contact during weeks $t-1$ or t; otherwise $e_{ijt} = 0$. Contacts are assumed to be symmetric (i.e., $e_{ijt} = e_{jit}$ for all i, j, t) and by convention $e_{iit} = 0$ for all i and t. The eX-FLU social contact network is represented graphically in Figure 1, where an edge between two nodes indicates that the two individuals reported contact in at least one week of the study, i.e., $\max_{t \in \{0,\ldots,\tau\}} e_{ijt} = 1$ for individuals i and j.

2.2 Potential outcomes with interference

Effects of the intervention can be represented by comparisons of potential outcomes (Neyman, 1923; Rubin, 2005). Let $y_{it}(\mathbf{a})$ denote the potential outcome that would have been observed for person i at time t if, possibly counter to fact, assignment $\mathbf{a}$ had been chosen. To begin, potential outcomes are assumed to be fixed, i.e., deterministic functions of $\mathbf{a}$. In Section 2.4 below, this assumption will be relaxed by allowing potential outcomes to be stochastic. Note that the notation $y_{it}(\mathbf{a})$ allows for interference to be completely general, i.e., the outcome of individual i may depend on the intervention assignment of any other individual. Therefore, in eX-FLU each individual has $\binom{h}{m}$ potential outcomes. By the causal consistency assumption (Cole and Frangakis, 2009), one of these potential outcomes $y_{it}(\mathbf{A}) = Y_{it}$ is observed and the remaining become counterfactual.

Inference about intervention effects generally requires some assumptions about the structure of the interference (Sävje et al., 2021). The commonly employed Stable Unit Treatment Value Assumption (SUTVA) supposes that there is no interference between units (Rubin, 1980). In the presence of interference, SUTVA is violated and other assumptions about the interference structure are needed for valid inference. The partial interference assumption posits that individuals can be partitioned into clusters such that there is no interference between clusters (Sobel, 2006). Even though individuals in the eX-FLU sample were assigned to intervention via cluster randomization, self-reported contacts between participants from different clusters suggests the partial interference assumption likely does not hold for this trial. Instead, the methods below rely on a weaker assumption that allows interference to occur between any two individuals who recently had contact. Specifically,

it is only assumed that there is not interference between individuals who have not had any recent contact.

2.3 Randomization-based tests

Randomization-based tests permit inference about experiments based on the exact (i.e., finite sample) distribution of a given test statistic under a specific hypothesis about the intervention effect (Rosenbaum, 2002a; Rosenberger et al., 2019). An RI test entails a null hypothesis H_0, a test statistic $T(\mathbf{a})$, and a method for determining the distribution of the test statistic under H_0. In addition to being a function of the randomization vector $\mathbf{a}$, the test statistic $T(\mathbf{a})$ is also a function of the observed data, which is left implicit for notational simplicity. The RI tests in this section consider the sharp null hypothesis of no intervention effect given by $H_0 : y_{it}(\mathbf{a}) - y_{it}(\mathbf{a}') = 0$ for all $i \in \{1, \ldots, n\}$ and $\mathbf{a}, \mathbf{a}' \in \mathcal{A}$ (Fisher, 1935). Rejecting this null hypothesis suggests that there is some intervention effect for at least one individual in the sample. Under H_0, $y_{it}(\mathbf{a}) = Y_{it}$ for all $\mathbf{a} \in \mathcal{A}$, which allows all unobserved potential outcomes $y_{it}(\mathbf{a})$ to be inferred from the observed data. The sharp null hypothesis is assessed by comparing the test statistic for the observed assignment, $T(\mathbf{A})$, to the distribution of test statistics for all possible treatment assignments under hypothesis H_0. Assuming values of the test statistic far from zero provide evidence against the null hypothesis, the two-sided p-value, $\rho_T(\mathbf{A}) = \sum_{\mathbf{a} \in \mathcal{A}} I(|T(\mathbf{A})| \leq |T(\mathbf{a})|)/|\mathcal{A}|$, is the proportion of the sampling distribution that is as or more extreme than the observed test statistic, where $I(\cdot)$ is the indicator function and for notational convenience $|\cdot|$ denotes either the absolute value of a number or the cardinality of a set. In settings where enumerating all possible elements of $\mathcal{A}$ is computationally prohibitive, Monte Carlo sampling can be used to approximate p-values (Rosenberger et al., 2019; Wang and Rosenberger, 2020; Wang et al., 2020).

An RI test can assess the presence of an intervention effect without any structural assumptions about the effect. Different possible test statistics are considered below. For example, the test statistic may equal an estimator of the average intervention effect under some working model. However, the validity of the RI framework for hypothesis testing does not rely on correct specification of the working model; that is, type I error control is guaranteed under the null even if the working model is mis-specified (Loh et al., 2020).

For instance, the test statistic could be based on the maximum likelihood estimator (MLE) of the parameters of a logistic regression model. One such working model is $\text{logit}\{E(Y_{it})\} = \beta_0 + \beta_1(\Sigma_j e_{ij,t-1}A_j)/(\Sigma_j e_{ij,t-1})$, for $t \in \{1, \ldots, \tau\}$. The parameter β_1 in this model quantifies the effect on the outcome at time t of the proportion of contacts at $t-1$ assigned to the intervention. In the eX-FLU trial, ILI status of peers at the previous time point is likely a relevant component of the intervention effect. Thus, another possible working model is $\text{logit}\{E(Y_{it})\} = \gamma_0 + \gamma_1(\Sigma_j e_{ij,t-1}Y_{j,t-1}A_j)/(\Sigma_j e_{ij,t-1}Y_{j,t-1})$, where γ_1 quantifies the effect of contacts with ILI being assigned to the intervention. These models can be used to calculate the Wald test statistics $T_1(\mathbf{A}) = \hat{\beta}_1/se_{\hat{\beta}_1}$ and $T_2(\mathbf{A}) = \hat{\gamma}_1/se_{\hat{\gamma}_1}$, where $\hat{\beta}_1$ and $\hat{\gamma}_1$ are the MLEs of β_1 and γ_1 respectively under the working logit model, $se_{\hat{\beta}_1}$ is the estimated standard error of $\hat{\beta}_1$ using the observed information, and $se_{\hat{\gamma}_1}$ is defined analogously. In their respective models, β_1 and γ_1 are zero if there is no effect.

The test statistic could also be based on an estimator for the average difference in the number of ILI infections that could be attributed to individuals in the intervention and control groups. Let $W_{it} = (\Sigma_j e_{ijt} Y_{jt}) / (\Sigma_j e_{ijt})$ be the proportion of person i's contacts at time t who have ILI. The average difference in these proportions by intervention assignment is $T_3(\mathbf{A}) = (\Sigma_{j,t} W_{jt} A_j) / (\Sigma_{j,t} A_j) - \{\Sigma_{j,t} W_{jt}(1-A_j)\} / \{\Sigma_{j,t}(1-A_j)\}$, where $\Sigma_{j,t}$ denotes $\sum_{j=1}^{n} \sum_{t=1}^{\tau}$. An analogous test statistic that accounts for which individuals reported ILI at the previous time point is $T_4(\mathbf{A}) = (\Sigma_{j,t} W_{jt} A_j Y_{j,t-1}) / (\Sigma_{j,t} A_j Y_{j,t-1}) - \{\Sigma_{j,t} W_{jt}(1 - A_j) Y_{j,t-1}\} / \{\Sigma_{j,t}(1 - A_j) Y_{j,t-1}\}$. Since statistics T_1 and T_3 use information from all participants, regardless of ILI status, these statistics may capture transmission from asymptomatic or unreported cases. However, asymptomatic individuals in the intervention group were not encouraged to socially distance themselves, suggesting T_1 and T_3 may have diminished power. On the other hand, statistics T_2 and T_4 utilize information from ILI cases to measure the effect of the intervention in contacts of individuals known to have ILI.

For longitudinal studies, the intervention effect can be assessed by considering all available time points simultaneously, as in the test statistics above. Alternatively, a pairwise analysis may be conducted that separately investigates each pair of consecutive time points. A pairwise approach may be preferred if the intervention effect changes over time. While the test statistics defined above incorporate information from all τ time points, analogous test statistics can be defined for a pairwise analysis.

2.4 Stochastic potential outcomes model

Randomization tests of the sharp null hypothesis provide information about the presence of an intervention effect, but do not provide information about the magnitude of such an effect. Point estimates and confidence intervals (CIs) of the intervention effect may also be desired. In the absence of interference, a common approach to constructing CIs entails inverting randomization tests. This approach typically relies on a constant treatment effect assumption (Wang and Rosenberger, 2020). For example, if each individual has two potential outcomes $y_i(0)$ and $y_i(1)$, then the additive treatment effect assumption supposes $y_i(0) = y_i(1) + \tau$ for all i and some fixed constant τ. Unfortunately, such treatment effect homogeneity assumptions are often unrealistic in many settings, especially when the outcome is binary and in the presence of interference (Rosenbaum, 2002a; Rigdon and Hudgens, 2015). Therefore, in this section, a stochastic potential outcomes model is considered that allows for heterogeneous treatment effects without assuming a constant treatment effect. Such a model will be used in the eX-FLU analysis in Section 4 to construct point estimates and CIs of the effect of the intervention on the risk of ILI transmission between social contacts.

Unlike the deterministic potential outcome model in Sections 2.2 and 2.3, here potential outcomes are considered stochastic (Robins, 1988; Robins and Greenland, 1989). In particular, the potential outcome for individual i at time t for assignment vector $\mathbf{a}$ may now vary probabilistically and is denoted by the random variable $Y_{it}(\mathbf{a})$, which may follow a distribution specific to each individual at each time point. Motivated by the eX-FLU trial, these stochastic potential outcomes are assumed to follow an ILI transmission probability model described below. The intervention effect is quantified by the model parameters, and

RI tests of non-sharp null hypotheses about the parameters are utilized to construct point estimates and CIs for the effect.

The stochastic potential outcome model is parameterized by ILI transmission probabilities from sources inside and outside of the observed network. Let p_1 denote the per-contact probability of ILI transmission from an individual assigned to the social distancing intervention to another individual. Similarly, define p_0 as the per-contact transmission probability from an individual not assigned to the intervention. Additionally, let ϵ represent the probability of transmission from outside of the observed network at each time point. The inclusion of ϵ in the model allows for the possibility that some eX-FLU participants may have been infected from individuals outside of the observed study network. Let $\theta = (p_0, p_1, \epsilon)$ with corresponding parameter space $\Theta = \{\theta : (p_0, p_1, \epsilon) \in [0, 1]^3\}$.

The transmission probability model can be tailored to the study design and subject matter knowledge. For instance, the model can incorporate dynamics specific to transmission of the outcome under study, such as pathogen-specific latency or infectious periods. Since ILI is caused by multiple viruses rather than a single pathogen, the transmission probability model used in the eX-FLU analysis assumes that individuals remain susceptible to ILI even after previous infections. Therefore, individuals with ILI at week $t - 1$ are considered eligible to get ILI anew in week t. In settings in which infection is known to confer immunity for some period of time, changes in susceptibility could be encoded into the model. Additionally, transmission is assumed to occur only between direct contacts.

The per-contact probability of transmission depends on multiple factors, including the social contact network, ILI status, and intervention assignment. Let B_{ijt} be the (unobserved) indicator of whether person i develops ILI at time t as a result of contact with person j. Assume $P(B_{ijt} = 1) = \pi_{ijt}(\mathbf{a})$ where $\pi_{ijt}(\mathbf{a}) = Y_{j,t-1}(\mathbf{a})e_{ij,t-1}\{a_j p_1 + (1-a_j)p_0\}$. Note that if individual j does not have ILI or individuals i and j are not social contacts at time $t-1$, then the probability that person j transmits ILI to person i is zero. Let O_{it} be the (unobserved) indicator that person i is infected from an individual outside of the observed network at time t, such that $P(O_{it} = 1) = \epsilon$.

Only one individual needs to successfully transmit ILI to individual i for person i to become infected, implying $Y_{it}(\mathbf{a}) = \max\{B_{i1t}, \ldots, B_{int}, O_{it}\}$. Assume $B_{i1t}, \ldots, B_{int}$ and O_{it} are mutually independent given the outcomes at the previous time point, such that

$$P\{Y_{it}(\mathbf{a}) = 1 \mid Y_{1,t-1}(\mathbf{a}), \ldots, Y_{n,t-1}(\mathbf{a})\} = r_{it}(\mathbf{a}), \tag{1}$$

where $r_{it}(\mathbf{a}) = 1 - (1 - \epsilon)\prod_{j=1}^{n}\{1 - \pi_{ijt}(\mathbf{a})\}$. Then the parameters θ can be estimated by maximum likelihood estimation, using the log likelihood function

$$l(\theta) = \sum_{t=1}^{\tau}\left[\sum_{i=1}^{n} Y_{it}\log\{r_{it}(\mathbf{A})\} + (1 - Y_{it})\log\{1 - r_{it}(\mathbf{A})\}\right], \tag{2}$$

under the Markov-type assumption that $\mathbf{Y}_t \perp\!\!\!\perp \{\mathbf{Y}_{t-2}, \ldots, \mathbf{Y}_1\} | \mathbf{Y}_{t-1}$ for all $t \in \{3, \ldots, \tau\}$. The MLE of θ does not appear to have a closed form, and therefore numerical optimization is used to find estimates $\hat{\theta} = (\hat{p}_0, \hat{p}_1, \hat{\epsilon})$ that maximize (2).

Confidence regions for θ may be constructed by inverting an RI test of $H_0 : \theta = \theta^*$ for $\theta^* = (p_0^*, p_1^*, \epsilon^*) \in \Theta$. The probability model parameters can be utilized in a test statistic that includes information from observed data as well as the hypothesis under consideration, such as $T_5(\mathbf{a}, \theta^*) = \|\hat{\theta} - \theta^*\|_2^2$ where $\|\cdot\|_2$ denotes the Euclidean norm. Since this H_0 is not a sharp hypothesis due to the randomness of the potential outcomes, the sampling distribution of the test statistic can no longer be constructed through enumeration for each $\mathbf{a} \in \mathcal{A}$. Instead, the sampling distribution is approximated as follows. First, a random sample of intervention assignments $\mathbf{a}$ is drawn from $\mathcal{A}$; denote this sample $\mathcal{A}_s$. Then for each $\mathbf{a} \in \mathcal{A}_s$, longitudinal ILI outcomes are stochastically generated according to (1) under the null $H_0 : \theta = \theta^*$, and the test statistic T_5 is evaluated. Then, the p-value $\rho_{T_5}(\mathbf{A}, \theta^*) = \sum_{\mathbf{a} \in \mathcal{A}_s} I\{T_5(\mathbf{A}, \theta^*) \leq T_5(\mathbf{a}, \theta^*)\} / |\mathcal{A}_s|$ is computed. Repeating this hypothesis testing process for all values of $\theta^* \in \Theta$, a $(1 - \alpha)$ confidence region for θ is then $\mathcal{C}_\theta = \{\theta^* \in \Theta : \rho_{T_5}(\mathbf{A}, \theta^*) > \alpha\}$.

The intervention effect can be defined by a contrast of transmission probabilities; the per-contact risk difference $\delta_\theta = p_1 - p_0$ is the focus below. A point estimate for δ_θ based on the MLE for θ is $\delta_{\hat{\theta}} = \hat{p}_1 - \hat{p}_0$, and a $(1 - \alpha)$ confidence region for δ_θ is $\mathcal{C}_\delta = \{\delta_{\theta^*} : \theta^* \in \mathcal{C}_\theta\}$. The minimum and maximum values of $\mathcal{C}_\delta$ form a $(1 - \alpha)$ CI for δ_θ. Determining the endpoints of this CI may be computationally challenging in practice. A computationally efficient stochastic search procedure for approximating the CI endpoints is described in the Appendix.

3. Empirical Results

3.1 Simulation study

The RI inferential methods presented above were evaluated via simulations designed to emulate the eX-FLU trial. Baseline networks were simulated with two different network models: an exponential random graph model (ERGM), and a scale-free model that allows for highly connected individuals or super-spreaders (Keeling and Eames, 2005). Simulation results from each model are provided separately below. To generate temporal variation in the network, one percent of contacts at time $t-1$ were randomly chosen to be non-contacts at time t, and one percent of non-contacts at time $t-1$ were randomly chosen to be contacts at time t for $t \in \{1, \ldots, \tau\}$.

To emulate the cluster randomization design of eX-FLU, each of $n = 504$ individuals was assigned to one of $h = 112$ clusters with roughly equal numbers of individuals per cluster. The intervention was assigned to $m = 56$ of these h clusters, resulting in a total of $\binom{112}{56} \approx 4 \times 10^{32}$ possible intervention assignments. Baseline ILI status, y_{i0}, was assigned such that 25 of the n individuals had ILI at time 0. Outcomes were sequentially generated according to (1), i.e., Y_{it} for $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, \tau\}$ was sampled from a Bernoulli

distribution with mean $r_{it}(\mathbf{A})$. Power of hypothesis tests, coverage of CIs, and bias of point estimates were computed for data simulated under various combinations of p_0, p_1, and ϵ.

R code for these simulations is available at github.com/shaina-alexandria/netrix.

3.2 Power of RI tests

The test statistics from Section 2.3 were evaluated empirically for power to detect the intervention effect δ_θ for values of θ in a region of Θ considered plausible for the eX-FLU trial. For each value of θ, 500 data sets were simulated. The power of each test was approximated by the proportion of datasets where the RI test p-values were less than or equal to $\alpha = 0.05$. For each hypothesis test, the sampling distributions of the test statistics were approximated via 1000 randomly chosen $\mathbf{a} \in \mathcal{A}$.

Power results are shown in Figure 2 for the scale-free model and different values of θ. The ERGM network power results were similar, and are provided in Web Figure 1. Statistics T_2 and T_4 were more powerful than T_1 and T_3, demonstrating that contacts without ILI at the previous time period were not informative about the intervention effect. Test statistics T_2 and T_1 also tended to be more powerful than T_4 and T_3, respectively, suggesting that utilizing a working logistic model may be preferable for assessing intervention effects in settings such as eX-FLU. All four statistics demonstrate type 1 error control, as is guaranteed by RI.

On the other hand, type 1 error control would not be expected in this setting if instead standard methods were employed that assume the observations are independently and identically distributed. To illustrate, the scale-free model simulations described above were repeated under the null hypothesis $\delta_\theta = 0$. For each simulated data set, whether to reject the null was determined by naively assuming the Wald statistics T_1 and T_2 follow standard Normal distributions. The results in Web Table 1 show that such a naive approach does not control the type 1 error, unlike the RI tests. Similar results were obtained when the simulations above were repeated using an ERGM network; see Web Table 2.

3.3 Coverage of confidence regions and bias

For the transmission probability model in Section 2.4, simulations were conducted to evaluate coverage of the 95% confidence regions for θ and confidence intervals for δ_θ under various combinations of the true data generating parameter θ. Separate simulations were conducted for each of the two network generation models. Confidence regions for θ were evaluated over a broad range of values in Θ to assess coverage for both plausible and extreme values of θ. Results for the ERGM network model in Web Figure 2 show the empirical coverage was generally close to the nominal level over the range of parameter values considered; results (not shown) for the scale-free model were similar.

Empirical coverage of the CI for δ_θ and bias of the estimator $\delta_{\hat{\theta}}$ were also evaluated for values of θ considered plausible in the eX-FLU trial. Separate simulation studies were conducted for the two network models. Power of the randomization test using statistic T_5 to reject the hypothesis $H_0 : \delta_\theta = 0$ was also evaluated. Results are presented in Web Table 3. For the scale-free model, the 95% CIs tended to have at least nominal coverage and $\delta_{\hat{\theta}}$ had small empirical bias. The test statistic T_5 had moderate power for θ values with larger

$|\delta_\theta|$ and controlled type 1 error, although power decreased as ϵ increased. Results from the ERGM model were similar.

4. The Spillover Effect of Social Distancing

In this section, data from the eX-FLU study is used to illustrate how the methods presented above can be used to draw inference about the effect of the social distancing intervention in the observed network of students. Baseline ILI information was collected in the initial week of the study, after randomization but before the intervention could plausibly spillover to others. Over the ten-week study period ($\tau = 9$), the eX-FLU trial captured three types of social contacts. Housing rosters provided roommate information, course rosters provided classmate information, and weekly questionnaires recorded self-reported contacts. Of the 590 total participants across 117 housing clusters, 522 had network information available via weekly self-reported contact surveys, housing information, or course rosters. Individuals with no contact information were omitted from the analysis.

Different contact types may experience different intervention effects. For instance, since the eX-FLU intervention asks students with ILI to isolate in their dorm room, the effect of the intervention on roommates and non-roommates may differ. The analysis presented here considers the non-roommate and classmate (NC) network, defined by all self-reported or classmate contacts that were not between roommates. The NC network included $n = 522$ participants from 115 clusters, with approximately nine contacts per week on average. The number of weekly contacts per individual ranged from zero to 46.

Since the primary results from the eX-FLU trial have not yet been published, the ILI outcome is not currently available for analysis. Therefore, ILI outcomes were simulated according to the transmission probability model in Section 2.4 based on the observed eX-FLU NC network data. The parameter value $\theta = (0.30, 0.15, 0.01)$ was chosen to illustrate the methods in a scenario that demonstrated moderate power in Section 3. Figure 3 shows the NC network by simulated ILI status during the study period.

Figure 4 displays the observed values of the test statistics T_1, T_2, T_3, and T_4 along with the corresponding sampling distributions and p-values for each test based on 1000 randomly sampled $a \in \mathcal{A}$. Test statistics T_2 and T_4 indicate strong evidence for an intervention effect ($\rho < 0.01$). Statistics T_1 and T_3 also suggest a possible intervention effect, although these p-values are larger. These results are consistent with the empirical finding in Section 3.2 that T_2 and T_4 tend to be more powerful than T_1 and T_3. Fitting the transmission probability model from Section 2.4 to the NC network yielded $\hat{\theta} = (0.30, 0.16, 0.01)$. Thus the estimated intervention effect was $\delta_{\hat{\theta}} = -0.14 (95\%\ \text{CI}\ -0.18, -0.10)$, close to the true value $\delta_\theta = -0.15$. These results demonstrate that the proposed methods can be used to detect and accurately quantify intervention effects in trials such as eX-FLU.

5. Discussion

Experiments conducted on social networks create an opportunity to study spillover effects. In such settings, RI methods are valid even in the presence of interference and non-random

sampling from a target population. In this article, RI methods were developed for analysis of the eX-FLU trial. Randomization tests were used to assess the null hypothesis of no intervention effect, and a stochastic potential outcomes transmission probability model was proposed to construct point estimates and confidence intervals of the magnitude of the intervention effect. The proposed methods allow for interference between individuals and do not assume constant treatment effects. While motivated by the eX-FLU trial, the transmission model may be tailored to other settings based on existing subject matter knowledge, such as information about latency periods or immunity for the disease pathogen of interest.

There are several other possible analyses of the eX-FLU trial. The study investigators collected additional contact information on a subset of participants via the iEpi smartphone application. The iEpi app uses Bluetooth location information to detect potential interactions between substudy participants. The app then sends prompts to the participants to collect information about the context surrounding the interaction. The contact information provided by iEpi is expected to be more accurate than the weekly self-reported contact questionnaires used in the above analysis. Future work may consider using iEpi contact information to improve inference about the eX-FLU intervention effect. Additionally, the transmission probability model considered here does not incorporate individual susceptibility characteristics, such as receipt of influenza vaccine or hand hygiene habits. Future research could incorporate individual-level covariates that may affect infection susceptibility. One such approach could entail randomization tests of residuals based on regression models of the outcome on baseline covariates under the null hypothesis of no treatment effect (Parhat et al., 2014). Often these residuals will have less variation than the outcomes, such that residual-based randomization tests can have greater power than tests that only utilize the outcome and ignore covariates (Rosenbaum, 2002b).

Randomization-based inference is appealing in that hypothesis testing is exact, i.e., type I error control is guaranteed. On the other hand, adequate statistical power is not guaranteed, and thus the choice of test statistic may be consequential. A test with lower power can fail to detect intervention effects when present, and if inverted can lead to large, uninformative confidence regions. Therefore it is important to select powerful test statistics in practice; as in Section 3.2, simulation studies based on the application at hand may inform test statistic selection (Bowers et al., 2013).

Likewise, the form of the stochastic potential outcome model assumed can also have a substantial impact on inferences drawn about a particular data set. Therefore in practice it is important to consider assessment of model fit and robustness of results to the assumed model. For certain test statistics, a randomization test can be viewed as simultaneously assessing the plausibility of both the null parameter value and the assumed model given the observed data (Bowers et al., 2016), and models which are mis-specified can result in empty confidence sets. In other words, the observed data may be incompatible with all possible parameter values of the assumed model; Loh et al. (2020) provide such an example when modeling the spillover effects of cholera vaccination. In such instances, empty confidence regions indicate lack of fit of the assumed model (Keele and Miratrix, 2019). See Bowers

et al. (2016) for additional considerations regarding goodness-of-fit tests in the context of interference and spillover effects.

As any model is an over-simplification to some extent, in practice considering multiple models may provide greater insight than inferences based on a single model (Rosenbaum, 2020). For each model considered, the methods in this paper can be used to determine regions of the parameter space, if any, that are compatible with the observed data. In the context of randomized trials within social networks, this approach allows investigators to characterize plausible intervention effects across different assumed models.

Increasingly, public health interventions are being designed to impact more of the population than is intervened on directly. The continued development of causal methods for inference about spillover effects of interventions within networks are therefore important to understanding public health and policy implications of interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank the the Editor, Associate Editor, an anonymous reviewer, and the Causal Inference with Interference working group in the Biostatistics Department at UNC-Chapel Hill (Bryan Blette, Kayla Kilpatrick, Taylor Krajewski, Sam Rosin, Bonnie Shook-Sa, Jaffer Zaidi) for their helpful suggestions. Dr. Aiello acknowledges funding by the Centers for Disease Control and Prevention (U01CK000185), National Institute of Health (R01 EB025021), and National Science Foundation (NSF 20–4354). This research was also supported by NIH grant R01 AI085073 and by a Gillings Innovation Laboratory award from the UNC Gillings School of Global Public Health. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

APPENDIX

Details of confidence interval construction

The following procedure is used to conduct a stochastic search for the minimum and maximum δ_θ values in $\mathscr{C}_\delta$. For computational efficiency, the procedure (i) limits the number of null hypotheses tested and (ii) approximates each p-value by randomly sampling 100 values of **a** from $\mathscr{A}$. Determining the confidence interval endpoints entails two steps.

The first step involves finding the values of θ^* that maximize and minimize the function $V(\theta^*) = (\delta_{\theta^*} - \delta_{\hat{\theta}})/[\{\rho_{T_5}(\mathbf{A}, \theta^*) - \alpha\}^2 + \nu]$, where ν is a small positive constant; for the results presented in this paper, $\nu = 0.0001$. Intuitively, the value of θ^* that maximizes $V(\theta^*)$ will have the largest value of δ_{θ^*} (such that the numerator of $V(\theta^*)$ is large) among values of θ^* where the corresponding p-value is close to α (such that the denominator of $V(\theta^*)$ is

small). Likewise, the minimizer of $V(\theta^*)$ will have the smallest value of δ_{θ^*} among the set of θ^* where the p-value is close to α. The nloptr non-linear optimizer function from the nloptr R package (Powell, 2006; Johnson, 2020) is used to search for the maximum and minimum of $V(\theta^*)$ in two separate function calls. Each θ^* and corresponding p-value $\rho_{T_5}(\mathbf{A}, \theta^*)$ calculated during the optimization process via nloptr are retained; let S be the set of these θ^* values.

In the second step, monotone splines are used to estimate the upper and lower bounds of $\mathscr{C}_\delta$. In particular, the upper bound of $\mathscr{C}_\delta$ is estimated as follows. Using only values of $\theta^* \in S$ such that $\delta_{\theta^*} \geq \delta_\theta$, a monotonic decreasing spline is fit with predictor δ_{θ^*} and outcome $\rho_{T_5}(\mathbf{A}, \theta^*)$ using the scam function in the scam R package (Pya, 2020). Let $f_U(\delta)$ be the fitted monotonic decreasing spline and define δ_U as the unique value of δ such that $f_U(\delta_U) = \alpha$. Analogously, let $f_L(\delta)$ be the monotonic increasing spline obtained using only values of $\theta^* \in S$ such that $\delta_{\theta^*} \leq \delta_\theta$, and let δ_L be the unique value of δ such that $f_L(\delta_L) = \alpha$. The $(1 - \alpha)$ CI for δ_θ is then (δ_L, δ_U).

This procedure for estimating the CI endpoints is illustrated using a single data set in Web Figure 3.

REFERENCES

Aiello AE, Simanek AM, Eisenberg MC, Walsh AR, Davis B, Volz E, et al. (2016). Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. Epidemics 15, 38–55. [PubMed: 27266848]

Aronow PM and Samii C (2017). Estimating average causal effects under general interference, with application to a social network experiment. Annals of Applied Statistics 11, 1912–1947.

Athey S and Imbens GW (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, pages 73–140. Elsevier, North-Holland.

Basse GW, Feller A, and Toulis P (2019). Randomization tests of causal effects under interference. Biometrika 106, 487–494.

Bowers J, Fredrickson M, and Panagopoulos C (2013). Reasoning about interference between units. Political Analysis 21, 97–124.

Bowers J, Fredrickson MM, and Aronow PM (2016). Research note: A more powerful test statistic for reasoning about interference between units. Political Analysis 24, 395–403.

Clemens J, Shin S, and Ali M (2011). New approaches to the assessment of vaccine herd protection in clinical trials. The Lancet Infectious Diseases 11, 482–487. [PubMed: 21616458]

Cole SR and Frangakis CE (2009). The consistency statement in causal inference: a definition or an assumption? Epidemiology 20, 3–5. [PubMed: 19234395]

Cox DR (1958). Planning of Experiments. John Wiley & Sons, Inc., New York.

Csardi G and Nepusz T (2006). The igraph software package for complex network research. InterJournal Complex Systems 1695, 1–9.

Eckles D, Karrer B, and Ugander J (2017). Design and analysis of experiments in networks: Reducing bias from interference. Journal of Causal Inference 5,.

Fisher RA (1935). The Design of Experiments. Oliver and Boyd, Edinburgh.

Forastiere L, Airoldi EM, and Mealli F (2021). Identification and estimation of treatment and interference effects in observational studies on networks. Journal of the American Statistical Association 116, 901–918.

Goel S, Salganik MJ, and Raftery A (2010). Assessing respondent-driven sampling. Proceedings of the National Academy of Sciences 107, 6743–6747.

Halloran E and Hudgens MG (2012). Causal inference for vaccine effects on infectiousness. The International Journal of Biostatistics 8, 1–40.

Hudgens MG and Halloran E (2008). Toward causal inference with interference. Journal of the American Statistical Association 103, 832–842. [PubMed: 19081744]

Johnson SG (2020). The NLopt nonlinear-optimization package. R package version 1.2.2.2.

Keele L and Miratrix L (2019). Randomization inference for outcomes with clumping at zero. The American Statistician 73, 141–150.

Keeling MJ and Eames KT (2005). Networks and epidemic models. Journal of the Royal Society Interface 2, 295–307. [PubMed: 16849187]

Klugman KP (2014). Herd protection induced by pneumococcal conjugate vaccine. The Lancet Global Health 2, e365–e366. [PubMed: 25103373]

Liu L, Hudgens MG, and Becker-Dreps S (2016). On inverse probability-weighted estimators in the presence of interference. Biometrika 103, 829–842. [PubMed: 29422692]

Loh WW, Hudgens MG, Clemens JD, Ali M, and Emch ME (2020). Randomization inference with general interference and censoring. Biometrics 76, 235–245. [PubMed: 31388990]

Neyman J (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. (Reprint 1990. Transl. by Dabrowska and Speed). Statistical Science 5, 465–472.

Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. (2020). Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database of Systematic Reviews 9, CD013574. [PubMed: 33959956]

Ogburn EL, Sofrygin O, Díaz I, and Van Der Laan MJ (2017). Causal inference for social network data. ArXiv e-prints.

Ogburn EL and Vanderweele TJ (2017). Vaccines, contagion, and social networks. The Annals of Applied Statistics 11, 919–948.

Parhat P, Rosenberger WF, and Diao G (2014). Conditional Monte Carlo randomization tests for regression models. Statistics in Medicine 33, 3078–3088. [PubMed: 24648378]

Paules CI, Marston HD, and Fauci AS (2020). Coronavirus infections - More than just the common cold. Journal of the American Medical Association 323, 707–708. [PubMed: 31971553]

Peak CM, Childs LM, Grad YH, Buckee CO, and Levin SA (2017). Comparing nonpharmaceutical interventions for containing emerging epidemics. Proceedings of the National Academy of Sciences 114, 4023–4028.

Powell MJ (2006). The NEWUOA software for unconstrained optimization without derivatives. In Large-scale Nonlinear Optimization, pages 255–297. Springer.

Pya N (2020). scam: Shape Constrained Additive Models. R package version 1.2–6.

Rigdon J and Hudgens MG (2015). Randomization inference for treatment effects on a binary outcome. Statistics in Medicine 34, 924–935. [PubMed: 25471299]

Robins J and Greenland S (1989). The probability of causation under a stochastic model for individual risk. Biometrics 45, 1125–1138. [PubMed: 2611320]

Robins JM (1988). Confidence intervals for causal parameters. Statistics in Medicine 7, 773–785. [PubMed: 3406605]

Rosenbaum P (2002a). Observational Studies. New York: Springer-Verlag.

Rosenbaum P (2020). Design of Observational Studies. Second Edition. Springer.

Rosenbaum PR (2002b). Covariance adjustment in randomized experiments and observational studies. Statistical Science 17, 286–327.

Rosenbaum PR (2007). Interference between units in randomized experiments. Journal of the American Statistical Association 102, 191–200.

Rosenberger WF, Uschner D, and Wang Y (2019). Randomization: The forgotten component of the randomized clinical trial. Statistics in Medicine 38, 1–12. [PubMed: 30047159]

Rubin DB (1980). Randomization analysis of experimental data: The Fisher randomization test. Journal of the American Statistical Association 75, 591–593.

Rubin DB (2005). Causal inference using potential outcomes. Journal of the American Statistical Association 100, 322–331.

Sävje F, Aronow PM, and Hudgens MG (2021). Average treatment effects in the presence of unknown interference. Annals of Statistics 49, 673 – 701. [PubMed: 34421150]

Sobel ME (2006). What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. Journal of the American Statistical Association 101, 1398–1407.

Tchetgen Tchetgen EJ, Fulcher I, and Shpitser I (2021). Auto-g-computation of causal effects on a network. Journal of the American Statistical Association 116, 833–844. [PubMed: 34366505]

Tchetgen Tchetgen EJ and VanderWeele TJ (2012). On causal inference in the presence of interference. Statistical Methods in Medical Research 21, 55–75. [PubMed: 21068053]

Wang Y and Rosenberger WF (2020). Randomization-based interval estimation in randomized clinical trials. Statistics in Medicine 39, 2843–2854. [PubMed: 32491198]

Wang Y, Rosenberger WF, and Uschner D (2020). Randomization tests for multiarmed randomized clinical trials. Statistics in Medicine 39, 494–509. [PubMed: 31846110]

Wu J and Ding P (2020). Randomization Tests for Weak Null Hypotheses in Randomized Experiments. Journal of the American Statistical Association, (in press).

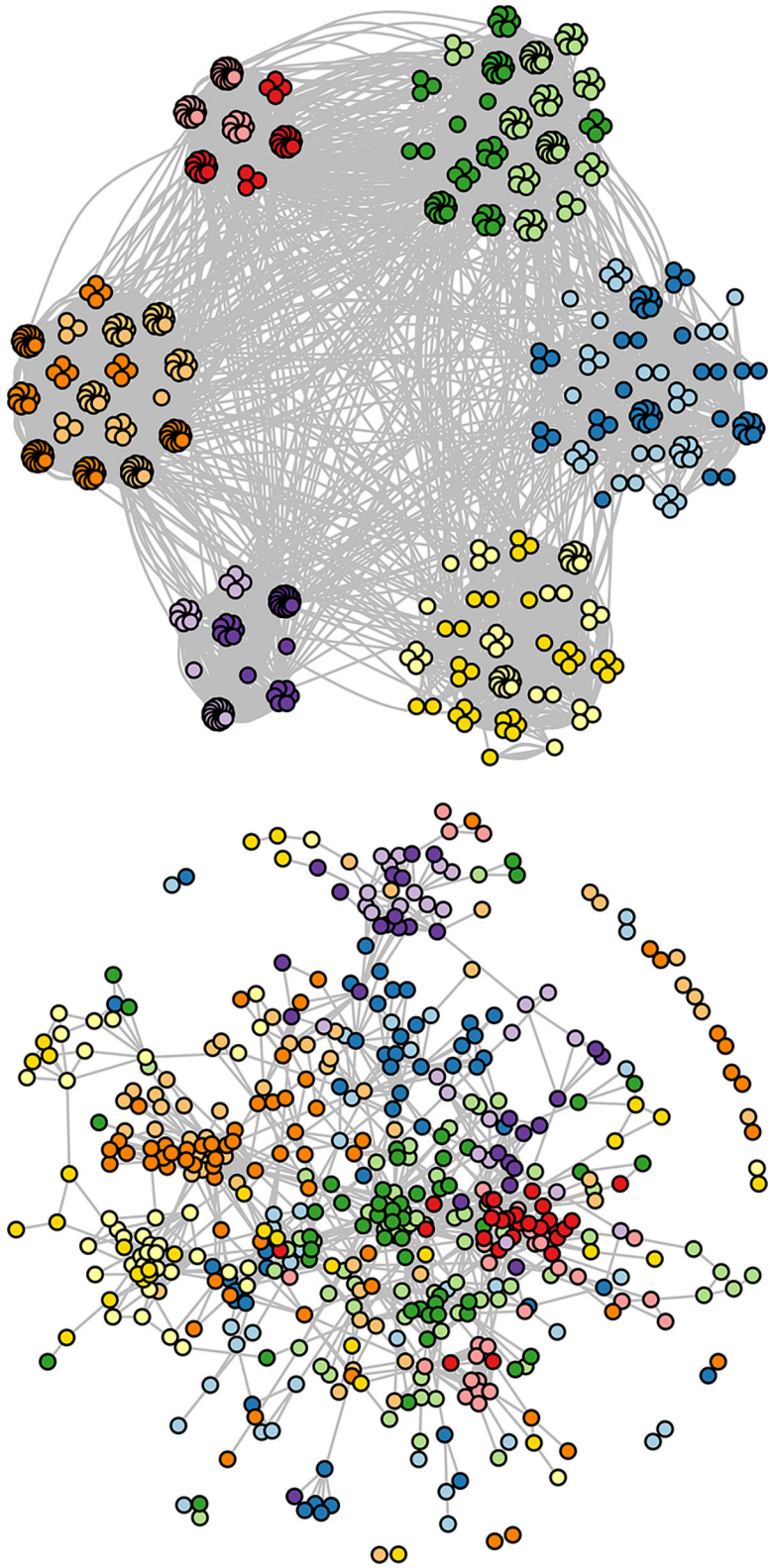


Figure 1:

Graphical representations of the eX-FLU randomization design. Each node represents one individual, and an edge between nodes indicates reported contact between individuals in at least one week of the study. Darker shades of each color represent cohort participants in the intervention group, and lighter shades represent participants in the control group. Figure (a) shows a random sample of 30% of the network edges in a layout where proximity between nodes is based on relative geographic location of participant residence halls, which were used as clusters for randomization of the intervention. Figure (b) shows 100% of network edges arranged in a layout that bases node proximity on frequency of reported contact using the igraph R package (Csardi and Nepusz, 2006). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

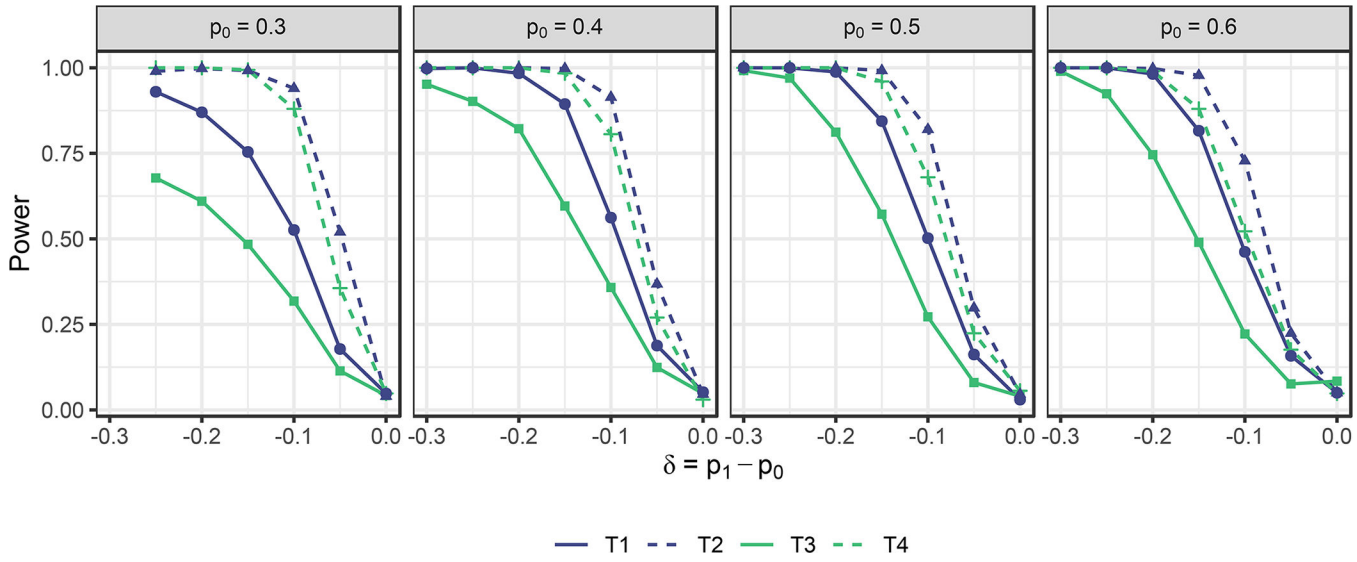
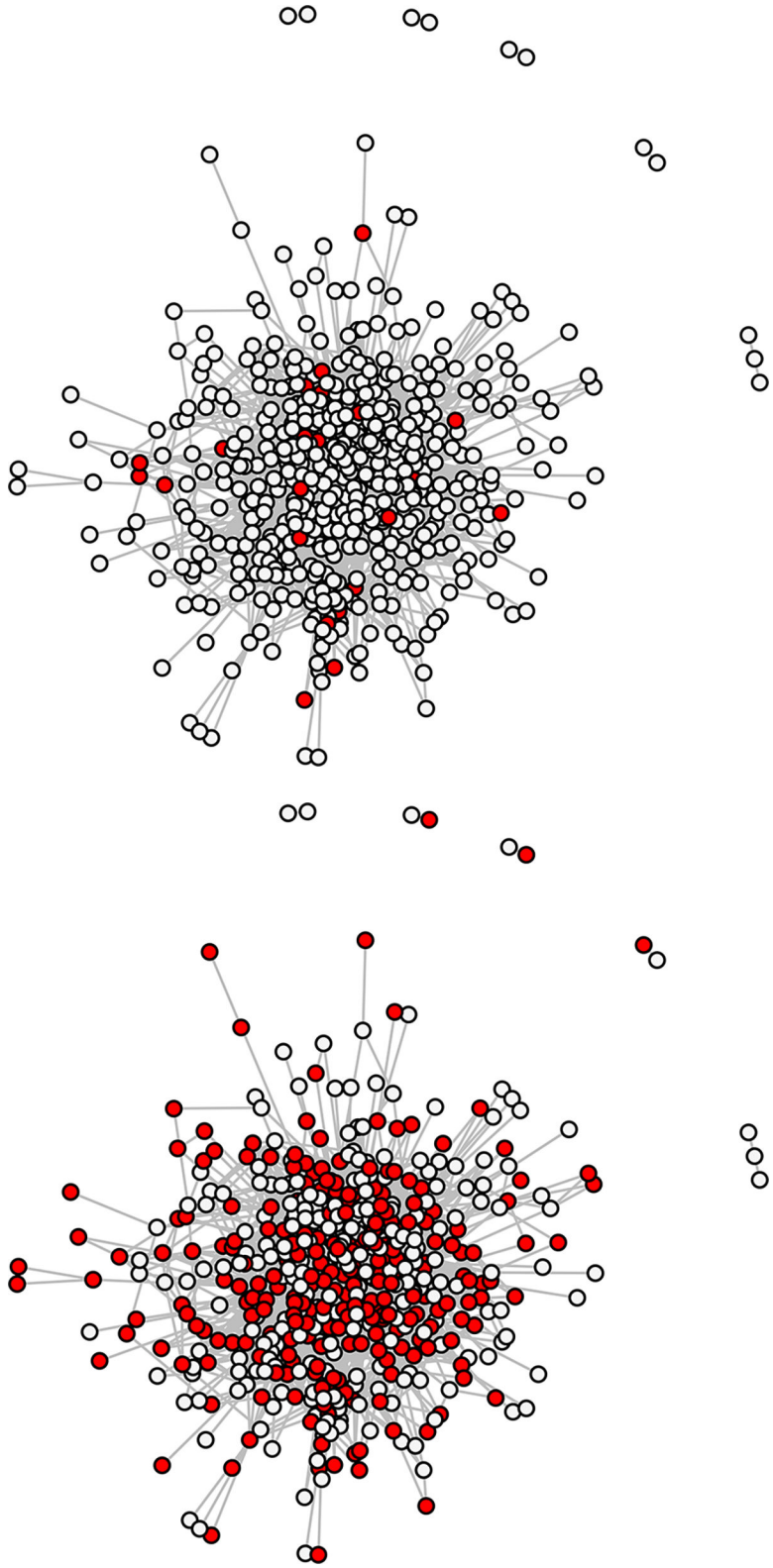


Figure 2:

Empirical power for test statistics T_1, T_2, T_3, and T_4 under the scale-free model with $n = 504$, $\tau = 9$, and 25 participants with ILI at baseline. Results are shown for $\epsilon = 0.01$ and various combinations of p_0 and p_1. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.



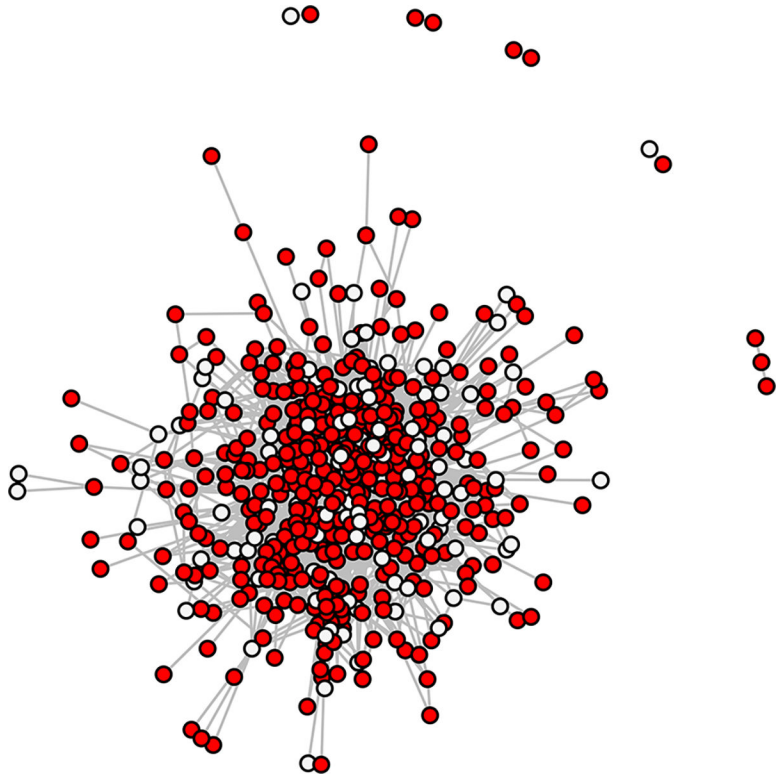


Figure 3:
Reported ILI in each of three weeks of the eX-FLU NC network. Nodes are colored red if the participant reported ILI during the specified week of the study period. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

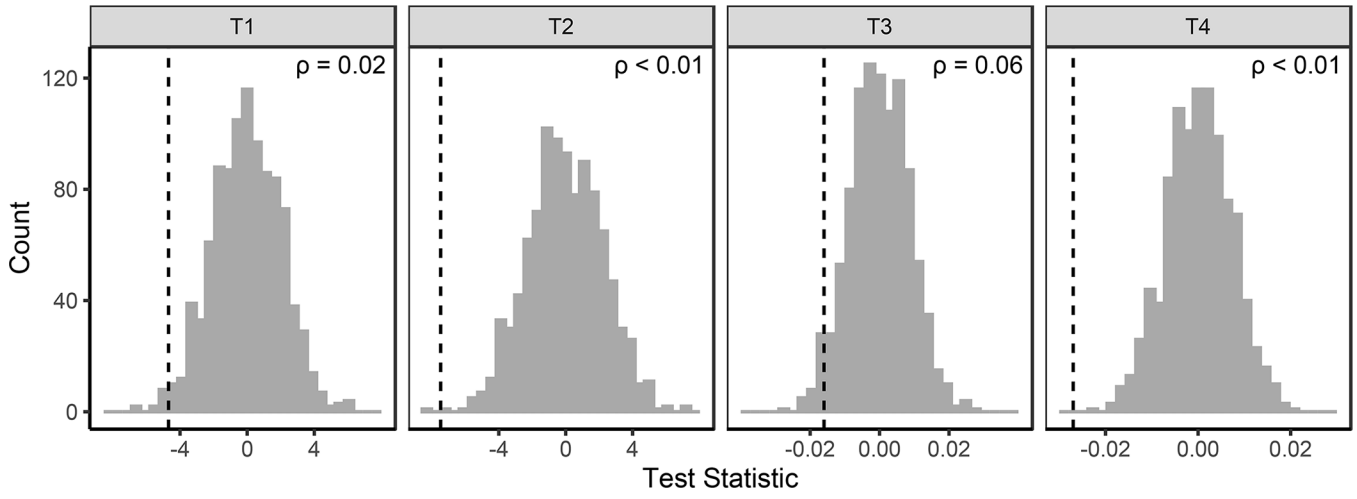


Figure 4:
Sampling distribution and RI p-value (ρ) for test statistics T_1, T_2, T_3, and T_4 based on the NC network in the eX-FLU trial. The observed test statistic is indicated by the vertical dashed line.