Causal Inference with Interference

Ashley L. Buchanan

Fall 2024

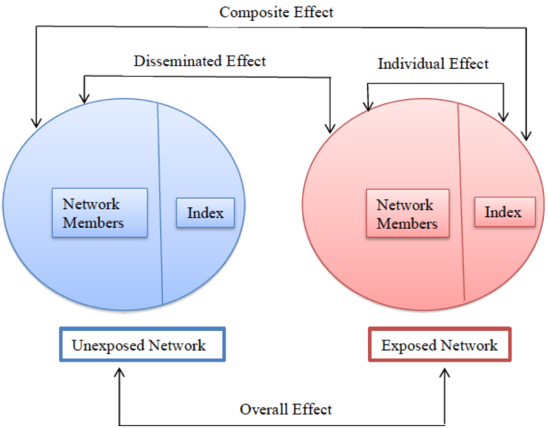
Outline

Introduction

- No Interference
  - Outcome of one individual assumed to be unaffected by the treatment assignment of others
  - Typical assumption of causal inference
  - Part of SUTVA
- Clearly not true in some settings
  - Infection diseases, education interventions. social sciences
  - Individuals often embedded in networks even if we ignore this in our study
- Phenomenon of interest vs. nuisance

Halloran and Struchiner (1991, 1995)

- The following slide shows different possible vaccine effects described by HS
- Several vaccine studies have been conducted or analyzed with the intent to estimate these effects (Moulton et al 2001; Longini et al 2002; Ali et al 2005; King et al 2006)
- Overlap in nomenclature with mediation literature (direct, indirect)


Composite Effect
Disseminated Effect
Individual Effect
Network Members
Index
Network Members
Index
Unexposed Network
Exposed Network
Overall Effect

Interference Definition

- That two potential outcomes sufficiently represent all potential outcomes for an individual assumes no interference between individuals
  i.e., the treatment of one individual does not affect the outcome of other individuals (Cox 1958)
- The no interference assumption may not hold in some settings
- Examples: Vaccine studies, educational intervention studies, HIV prevention studies
- Settings: Epidemiology, medical research, econometrics, social network analysis

SUTVA (Fine Points 1.1-1.2)

- No interference is part of SUTVA (Rubin 1980)
- Stable Unit Treatment Value Assumption
    1. No interference
    2. Only one version of treatment and one version of no treatment (control)

Or if there are multiple versions of treatment, they are irrelevant (Cole and Frangakis 2008, VanderWeele 2009, Pearl 2010)

See Fine Point 1.2 regarding multiple versions of treatment

General Approach

- Population of groups of individuals (blocks of units; clusters)
- Assume partial interference: Possibly inteference between individuals in a group but not between groups.
- Define direct, disseminated (indirect), composite (total) causal effects
- Two-stage randomization
    1. Groups to allocation strategies α_1, α_0
    2. Given 1, individuals randomized to treatment/controls $A \in 0, 1$
- Unbiased estimators, variance using randomization-based inference or M-estimation

Example: Vaccine Trial

- Groups: Schools sufficiently separated geographically
- Individuals: Students
- Assignment mechanism
  1. Randomized some schools to 50%, others to 25% vaccine coverage
  2. Randomized students to vaccine or placebo conditional on school assignment strategy from step 1

Notation

- N groups; n_i individuals in groups $i = 1, \ldots, N$
- $\mathbf{A}_i = (A_{1i}, \ldots, A_{1n_i})$ treatments received for n_i individuals in group i
  $A_{ij} = 0$ or 1 implies $\mathbf{A}_i$ can take on 2^{n_i} possible values
  $\mathbf{A}_{i,-j}$ is the $n_i - 1$ subvector of $\mathbf{A}_i$ with the j^{th} entry deleted
  $\mathbf{a}_i$ and a_{ij} denote possible values of $\mathbf{A}_i$ and A_{ij}
- Let $A(n)$ be the set of vectors of all possible exposure allocations of length n. e.g., $A(2) = \{(0,0), (0,1), (1,0), (1,1)\}$, $\mathbf{a}_i \in R^{n_i}$
- $A(n, k)$ denotes when exactly k individuals receive treatment 1 (i.e., completely randomized design)
- Let α be the proportion assigned to treatment in a group

Assignment Mechanism

- $S_i = 1$ if the i^{th} group is assigned to α_1 and 0 otherwise
  $\mathbf{S} = (S_1, \ldots, S_N)$
  $C = \sum_i S_i$
- Parameterization for treatment assignment strategy
  - Complete randomized group assignment strategy if k_i number treated in block i , i.e., $\pi(a_i, \alpha) = I(a_i \in A(n_i, k_i))/\binom{n_i}{k_i}$
  - Bernoulli Allocation: $\pi(\mathbf{a}_i, \alpha) = \prod_{j=1}^{n_i} \alpha^{a_{ij}}(1-\alpha)^{1-a_{ij}}$

Potential Outcomes

- $y_{ij}(\mathbf{a}_i)$ is the potential outcome of individual j in group i under $\mathbf{a}_i$
- Allows for interference between individuals within group i
- Can write $y_{ij}(\mathbf{a}_i)$ as $y_{ij}(\mathbf{a}_{i,-j}, a_{ij} = a)$
- Have 2^{n_i} potential outcomes per individual, instead of 2 potential outcomes per individual in the absence of interference

Average Potential Outcomes

- Individual average potential outcome

$$\bar{y}_{ij}(a,\alpha) = \sum_{a_{i,-j} \in A(n_i - 1)} y_{ij}(\alpha, a_{ij} = a) \Pr(\mathbf{A}_{i,-j} = a_{i,-j} | A_{ij} = a)$$

- Group average potential outcome

$$\bar{y}_i(a,\alpha) = \frac{1}{n_i} \sum_{i=1}^{n_i} \bar{y}_{ij}(a,\alpha)$$

- Population average potential outcome

$$\bar{y}(a,\alpha) = \frac{1}{N} \sum_{j=1}^{N} \bar{y}_i(a,\alpha)$$

Average Potential Outcomes: Example

- Suppose that group 1 has the following potential outcomes $y_{1j}(\mathbf{x}_1)$

j	000	001	010	100	011	101	110	111
1	1	2	3	4	5	6	7	8
2	9	10	11	12	13	14	15	16
3	17	18	19	20	21	22	23	24

- Suppose completely randomized individual treatment assignment with $K_1 = 2$ for α_1 and $K_1 = 1$ for α_0

$$\bar{y}_1(0, \alpha_1) = ?$$
$$\bar{y}_1(0, \alpha_0) = ?$$

Average Potential Outcomes: Example

- Suppose that group 1 has the following potential outcomes $y_{1j}(\mathbf{x}_1)$

j	000	001	010	100	011	101	110	111
1	1	2	3	4	5	6	7	8
2	9	10	11	12	13	14	15	16
3	17	18	19	20	21	22	23	24

- Suppose completely randomized individual treatment assignment with $K_1 = 2$ for α_1 and $K_1 = 1$ for α_0

$$\bar{y}_1(0, \alpha_1) = \frac{5 + 14 + 23}{3} = 14$$

$$\bar{y}_1(0, \alpha_0) = \frac{(2+3)/2 + (10+12)/2 + (19+20)/2}{3} = 11$$

Average Potential Outcomes

- Marginal individual average potential outcomes

$$\bar{y}_{ij}(\alpha) = \sum_{a_i \in A(n_i)} y_{ij}(\alpha_1) \Pr(\mathbf{A}_i = \mathbf{a}_i)$$

- Marginal group and population average potential outcomes

$$\bar{y}_i(\alpha) = \frac{1}{n_i} \sum_{i=1}^{n_i} \bar{y}_{ij}(\alpha)$$

$$\bar{y}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \bar{y}_i(\alpha)$$

Causal Estimands: Direct Effects

- Individual direct causal effect of treatment 0 compared to treatment 1 for the individual j in group i by

$$CE_{ij}^{D}(\alpha) = y_{ij}(a_{ij} = 1, \alpha) - y_{ij}(a_{ij} = 0, \alpha)$$

- Individual average direct causal effect

$$\overline{CE}_{ij}^{D}(\alpha) = \bar{y}_{ij}(1, \alpha) - \bar{y}_{ij}(0, \alpha)$$

- Group average direct causal effect

$$\overline{CE}_{i}^{D}(\alpha) = \bar{y}_{i}(1, \alpha) - \bar{y}_{i}(0, \alpha)$$

- Population average direct causal effect

$$\overline{CE}^{D}(\alpha) = \bar{y}(1, \alpha) - \bar{y}(0, \alpha)$$

Q: Which of the quantities above can never be identified in the observed data (A, Y)?

Causal Estimands: Indirect Effects

- Individual indirect causal effect of treatment programs α_1 compared with α_1 on individual j in group i by

$$CE^I_{ij}(\alpha_1, \alpha_0) = y_{ij}(\alpha_1, a_{ij} = 0) - y_{ij}(\alpha_0, a_{ij} = 0)$$

- Individual average indirect causal effect

$$\overline{CE}^I_{ij}(\alpha_1, \alpha_0) = \bar{y}_{ij}(0, \alpha_1) - \bar{y}_{ij}(0, \alpha_0)$$

- Group average indirect causal effect

$$\overline{CE}^I_{i}(\alpha_1, \alpha_0) = \bar{y}_{i}(0, \alpha_1) - \bar{y}_{i}(0, \alpha_1)$$

- Population average indirect causal effect

$$\overline{CE}^I(\alpha_1, \alpha_0) = \bar{y}(0, \alpha_1) - \bar{y}(0, \alpha_0)$$

Causal Estimands: Total and Overall Effects

- Population average total causal effect

$$\overline{CE}^{T}(\alpha_1, \alpha_0) = \bar{y}(1, \alpha_1) - \bar{y}(0, \alpha_0)$$

- Population average overall causal effect

$$\overline{CE}^{O}(\alpha_1, \alpha_0) = \bar{y}(\alpha_1) - \bar{y}(\alpha_0)$$

Causal Estimands: Remarks

- Total = direct + indirect
- Estimands in general depend on treatment allocation strategy
- Under no interference

$$y_{ij}(\mathbf{a}_i) = y_{ij}(\mathbf{a}_i') \text{ for all } \mathbf{a}_i, \mathbf{a}_i' \text{ such that } a_{ij} = a_{ij}'$$

  - Indirect causal effects are zero
  - Total causal effect equals direct causal effect
  - Causal effects not dependent on the treatment strategies

Estimators

- Assumption 1: Completely randomized assignment strategy
- For $a = 0, 1$:

$$\widehat{Y}_i(a, \alpha) = \frac{\sum_j Y_{ij} I(A_{ij} = a)}{\sum_j I(A_{ij} = a)} = \frac{1}{n_i} \sum_j \frac{Y_{ij} I(A_{ij} = a)}{\Pr[A_{ij} = a | S_i = 1]}$$

Under assumption 1, $E[\widehat{Y}_i(a, \alpha) | S_i = 1] = \bar{y}_i(a, \alpha)$.

Example Revisited

- Suppose that group 1 has the following potential outcomes $y_{1j}(\mathbf{x}_1)$

j	000	001	010	100	011	101	110	111
1	1	2	3	4	5	6	7	8
2	9	10	11	12	13	14	15	16
3	17	18	19	20	21	22	23	24

- Under Assumption 1, with $K_1 = 2$ for α_1 and $K_1 = 1$ for α_0

$$E\{\widehat{Y}_i(0, \alpha_1)|S_i = 1\} = ?$$

$$E\{\widehat{Y}_i(0, \alpha_0)|S_i = 0\} = ?$$

Example Revisited

- Suppose that group 1 has the following potential outcomes $y_{1j}(\mathbf{x}_1)$

j	000	001	010	100	011	101	110	111
1	1	2	3	4	5	6	7	8
2	9	10	11	12	13	14	15	16
3	17	18	19	20	21	22	23	24

- Under Assumption 1, with $K_1 = 2$ for α_1 and $K_1 = 1$ for α_0

$$E\{\hat{Y}_1(0,\alpha_1)|S_1 = 1\} = \frac{5+14+23}{3} = 14 = \bar{y}_1(0,\alpha_1)$$

$$E\{\hat{Y}_1(0,\alpha_0)|S_1 = 0\} = \frac{1}{3}\{\frac{2+10}{2} + \frac{3+19}{2} + \frac{12+20}{2}\} = 11 = \bar{y}_1(0,\alpha_0)$$

Estimators

- $\widehat{CE}^{D}(\alpha_1) = \widehat{Y}(1, \alpha_1) - \widehat{Y}(1, \alpha_1)$
- $\widehat{CE}^{I}(\alpha_1, \alpha_0) = \widehat{Y}(0, \alpha_1) - \widehat{Y}(0, \alpha_0)$
- $\widehat{CE}^{T}(\alpha_1, \alpha_0) = \widehat{Y}(1, \alpha_1) - \widehat{Y}(0, \alpha_0)$

Overall Estimators

- $\widehat{Y}_i(\alpha) = \frac{\sum_j Y_{ij}}{n_i}$
- $\widehat{Y}(\alpha) = \frac{\sum_i \hat{Y}_i(\alpha) I[S_i=1]}{\sum_i I[S_i=1]}$
- Under assumption 1, $E\{\widehat{Y}(\alpha)\} = \bar{y}(\alpha)$
- Unbiased estimator: $\widehat{CE}^O(\alpha_1, \alpha_0) = \widehat{Y}(\alpha_1) - \widehat{Y}(\alpha_0)$

Variance

- Unbiased estimators of the variance of the estimators does not exist without further assumptions
- Stratified Interference (SI): Only matters how many were treated in group or cluster, and does not matter who was treated
- For a given $a_{ij} = a$, individual j in group i has
  1 potential outcome assuming no interference
  n_i potential outcomes assuming stratified interference
  $2^{n_i - 1}$ potential outcomes under no assumptions
- Under SI, simple random sampling and two stage cluster sampling yield unbiased estimators of variance of $\hat{Y}_i(0, \alpha)$ and $\hat{Y}(0, \alpha_1)$
- Variance estimators are unbiased when effect is additive, positively biased otherwise

Illustrative Example

Two-stage randomized placebo-controlled vaccine trial based on data from Ali et al. (2005)

	α	Vaccine ($X_{ij} = 1$)		Placebo ($X_{ij} = 0$)	
		Total	Cases	Total	Cases
		$\sum_j X_{ij}$	$\sum_j X_{ij} Y_{ij}$	$\sum_j (1 - X_{ij})$	$\sum_j (1 - X_{ij}) Y_{ij}$
1	α_1	12541	16	12541	18
2	α_1	11513	26	11513	54
3	α_1	10772	17	25134	119
4	α_0	8883	22	20727	122
5	α_0	5627	15	13130	92

α_0 is the allocation strategy for the group that randomized 50% to the treatment.
α_1 is the allocation strategy for the group that randomized 30% to the treatment.

Estimates of population average effects per 1000 individuals per year

Effect	Parameter	Estimate	Estimated Variance
Direct	$\overline{CE}^{D}(\alpha_1)$	1.30	0.856
Direct	$\overline{CE}^{D}(\alpha_0)$	3.64	0.178
Indirect	$\overline{CE}^{I}(\alpha_1, \alpha_0)$	2.81	3.079
Total	$\overline{CE}^{T}(\alpha_1, \alpha_0)$	4.11	0.672
Overall	$\overline{CE}^{O}(\alpha_1, \alpha_0)$	2.37	1.430

- Indirect: 50% vaccine coverage results in 2.8 fewer cholera cases per 1000 unvaccinated individuals per year compared to 30% vaccine coverage
- Overall: 50% vaccine coverage results in 2.4 fewer cholera cases per 1000 individuals per year compared to 30% vaccine coverage

Observational Studies

- Methods in the presence of interference often rely on randomization and the assumption of partial interference, but provides a solution to the problem of interference in randomized and nonrandomized designs with nonoverlapping clusters (e.g. Sobel, 2006; Hong and Raudenbush, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2012; Liu and Hudgens, 2014; Buchanan et al, 2018).
- Suppose a two-stage randomization not employed, but instead we have an observational study
- Tchetgen Tchetgen and VanderWeele (2012) suggest IPW estimator where all observations from group i are weighted by the inverse of probability of the treatment assignment vector $\mathbf{A}_i$ given $\mathbf{X}_i$
- Essentially, standardizing to a counterfactual study with a Bernoulli allocation mechanism
- Alternative approaches to standardize to a study with correlation between the treatment assignment mechanisms (Barkley, et al 2020; Papadogeorgou et al. 2019)

Parameters

- $\pi(a_{k,-i};\alpha) = \Pr(A_{k,-i} = a_{k,-i}) = \prod_{j=1,j\neq i}^{n_k} \alpha^{a_{kj}}(1-\alpha)^{1-a_{kj}}$
- $\pi(a_k;\alpha) = \Pr(A_k = a_k) = \prod_{j=1}^{n_k} \alpha^{a_{kj}}(1-\alpha)^{1-a_{kj}}$.
- $\bar{y}_{ki}(a,\alpha) = \sum_{a_{k,-i}} y_{ki}(a_i = a, a_{k,-i})\pi(a_{k,-i};\alpha)$.
- Averaging over all individuals in each cluster, then over all clusters, we define the population average potential outcome as $\bar{y}(a,\alpha) = \sum_{k=1}^{K}\{\sum_{i=1}^{n_k} \bar{y}_{ki}(a,\alpha)/n_k\}/K$.
- Define the marginal average potential outcome for individual i under allocation strategy α by $\bar{y}_{ki}(\alpha) = \sum_{a_k} y_{ki}(a_k)\pi(a_k;\alpha)$.
- Averaging over individuals within each cluster, then over all clusters, define the population average potential outcome as $\bar{y}(\alpha) = \sum_{k=1}^{K}\{\sum_{i=1}^{n_k} \bar{y}_{ki}(a)/n_k\}/K$.
- Effects can be defined such as the spillover or indirect $\overline{IE}(\alpha,\alpha') = \bar{y}(0,\alpha) - \bar{y}(0,\alpha')$

Assumptions

- Representativeness of the unexposed for the treatment response had they been exposed and vice versa conditional on baseline covariates (i.e., conditional exchangeability at the cluster level)
  $\Pr(\mathbf{A}_i = \mathbf{a}_i | \mathbf{L}_i, \mathbf{Y}_i(\cdot)) = \Pr(\mathbf{A}_i = \mathbf{a}_i | \mathbf{L}_i)$
- Homogeneity of treatment effects despite any variations that may occur in practice, and no multiple versions of treatment
- No measurement error in any variable needed for valid analysis
- No interference between clusters.
- Cluster-level positivity assumption for the propensity score.
  $\Pr(\mathbf{A}_i = \mathbf{a}_i | \mathbf{L}_i) > 0$

Estimators

Cluster-level propensity score can be calculated by adjusting with individual level covariates among those in the cluster.

$$f_{A_i|X_i}(A_i|X_i;\theta_x,\theta_s) = \int \prod_{j=1}^{n_i} h_{ij}(b_i;\theta_x)^{A_{ij}}\{1 - h_{ij}(b_i;\theta_x)\}^{1-A_{ij}} f_b(b_i;\theta_s)db_i$$

where $h_{ij}(b_i;\theta_x) = Pr(a_{ij} = 1|X_{ij}, b_i, \theta_x) = logit^{-1}(X_{ij}\theta_x + b_i)$ is a propensity score for individual j in cluster i and $f_b(\cdot;\theta_s)$ is the density of cluster specific random effect $b_i \sim N(0,\theta_s)$.

IPW estimation: IPW estimator

IPW estimator for group-level potential outcome:

$$\widehat{Y}_i^{ipw}(a, \alpha) = \frac{\sum_{j=1}^{n_i} \pi_i(A_{i,-j}; \alpha) I(A_{ij} = \alpha) Y_{ij}}{n_i f_{A_i|X_i}(A_i|X_i; \hat{\theta})}$$

Marginal potential outcome:

$$\widehat{Y}_i^{ipw}(\alpha) = \frac{\sum_{j=1}^{n_i} \pi_i(A_i; \alpha) I(A_i = \alpha) Y_{ij}}{n_i f_{A_i|X_i}(A_i|X_i; \hat{\theta})}$$

Population-level IPW estimators

$$\widehat{DE}(\alpha) = \widehat{Y}^{ipw}(a=0;\alpha) - \widehat{Y}^{ipw}(a=1;\alpha)$$

$$\widehat{IE}(\alpha,\alpha') = \widehat{Y}^{ipw}(a=0;\alpha) - \widehat{Y}^{ipw}(a=0;\alpha')$$

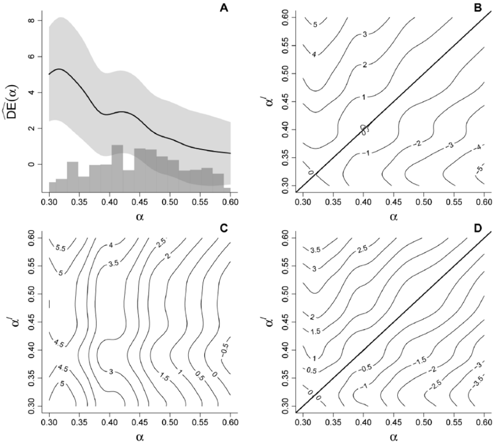
$$\widehat{TE}(\alpha,\alpha') = \widehat{Y}^{ipw}(a=0;\alpha) - \widehat{Y}^{ipw}(a=1;\alpha')$$

$$\widehat{OE}(\alpha,\alpha') = \widehat{Y}^{ipw}(\alpha) - \widehat{Y}^{ipw}(\alpha')$$

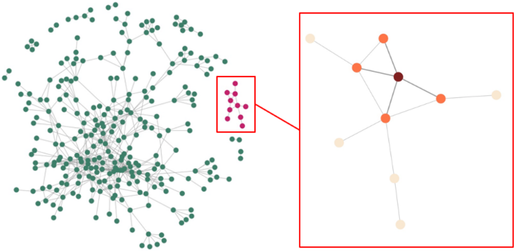
coverage: $\alpha < \alpha'$

Illustrative Example (Perez-Heydrich et al, 2014)

Individually randomized placebo-controlled vaccine trial based on data from Clemens et al. (1988) with (A) direct; (B) indirect; (C) total; and (D) overall.



Introduction to Networks



- Each node (person) has an outcome, treatment and covariates (attributes)
- Nodes are connected through edges, which represent social, work, school, sexual, healthcare, drug use/injection drug use, etc. partnerships
- Estimands: peer effects, treatment effects, spillover/interference effects, effects of network interventions
- Challenges:
  1. How to define and identify causal effects in a network-based study
  2. How to quantify uncertainty with complex network dependence

Approaches in the Literature

- Christakis and Fowler (2007, 2008, 2009, 2010, 2011, 2012) estimated peer effects in social network data
  - Model: $Y^{t}_{ego} \sim Y^{t-1}_{alter}, Y^{t-2}_{alter}, Y^{t-2}_{ego}, \mathbf{C}_{ego}$
  - Results included significant peer effects for obesity, smoking, alcohol consumption, etc.
  - Peer effects evaluated in other settings (Ali and Dwyer, 2009, Cacioppo et al, 2009; 2008; Lazer et al., 2010; Rosenquist et al, 2010, Wasserman, 2012)

Randomization-based Inference

- Randomization-based inference for networks (e.g., Toulis and Kao, 2013; Bowers et al., 2013; Aronow and Samii, 2013; Eckles et al., 2014, Choi 2016).
  - Assumes on finite population of N individuals and for each individual there is a set of individuals that may interfere with that individual (i.e., interference sets, neighborhoods, friends)
  - Interference sets can be represented by an adjacency matrix and often assumed to be known and fixed

Non-randomized Interventions

- In many studies, the intervention or treatment is not randomized
- There may be confounding at either the individual, network-level or both
- Also face issues of network dependence and homophily
- Complex dependencies between observations
- Current methods employ
  - A generalized propensity score (Forastiere, 2020) or a Bayesian generalized propensity score (Forastiere, 2018) that account for individual and neighborhood covariates
  - Targeted maximum likelihood estimation (TMLE) (Sofrygin, 2015)

Statistical Dependence in Networks

- Latent variables (i.e., homophily) lead to similar outcomes among close contacts
- Networks often observed at a single time point, so difficult to disentangle homophily from an effect
- Why is this a problem?
  - We cannot assume independence (i.e., cannot assume independent and identically distributed (iid))
  - Central limit theorem may not hold
  - Standard error estimates and confidence intervals will be anti-conservative!
- Network dependence (e.g., autocorrelation) is another threat to validity (particularly for single site studies) that can create bias (different from confounding and homophily!)

Possible Solutions for Dependence in Networks

- Create conditionally independent units; analyze with standard models, but conditional on information barriers (Ogburn and Vanderweele, 2017)
- Extension of influence function from iid setting with interference set (van der Laan, 2014) and social network setting with contagion and homophily (Ogburn, et al., 2017)
- Nearest neighbor approach: Potential outcomes of any individual only depends their own exposure and on exposures of their nearest neighbors (or two-step neighbors) (Lee, et al, 2021; Forastiere, et al., 2020)
- Subsampling: Implementation and conditions may not be applicable to networks (e.g., bootstrap)
- K-dependence: $Cov(W_i, W_j) = \sigma_k$, where $k = ||i, j||$ and estimate using a plug-in estimator

Proposed Methodology

Nearest neighborhood IPW estimator

We propose an inverse probability weighted (IPW) estimator where the interference set is defined as the set of the individual's nearest neighbors within the network.

IPCW estimator

The nearest neighborhood IPW estimator was extended into a setting with missing outcomes using inverse probability censoring weights (IPCW), where we consider two different assumptions for the censoring mechanism:

(1) censoring indicators are independent across participants

(2) censoring indicators are correlated between participants within a connected subnetwork or component in the social network.

Motivating Study: Transmission Reduction Intervention Project (TRIP)

- Sociometric network-based study of injection drug users in Athens, Greece from 2013 to 2016.
- Intervention: **community alerts**.
  Outcome: **the HIV risk behavior at 6-month follow up**.

Baseline interview

- 277 participants
- 29 alerted

$\xrightarrow{\text{six-month time period}}$

follow up interview

- 57 missing
- 4 were alerted

Figure 1: TRIP network with isolates removed. There are 277 participants and 542 links in the network.

Notations

Let $i = 1, 2, \cdots, n$ denote each participant in the study.

- A_i: the self-selected binary treatment/exposure of participant i
- Z_i: the vector of covariates for participant i
- $\mathcal{N}_i$: the set of participants that share a link with i
- d_i: $|\mathcal{N}_i|$, the degree of node i
- $A_{\mathcal{N}_i}$: the vector of baseline exposures for participants in $\mathcal{N}_i$
- $Z_{\mathcal{N}_i}$: the vector of baseline covariates for participants in $\mathcal{N}_i$
- C_i: the binary censoring indicator for participant i. i.e. due to loss to follow-up or administrative end of the study.

Potential Outcomes

- We assume Bernoulli counterfactual treatment allocation strategy with coverage α ($\sim$ participants in $\mathcal{N}_i$ are exposed with prob. α).
- Let $\pi(a_{\mathcal{N}_i};\alpha) = \alpha^{\sum a_{\mathcal{N}_i}}(1-\alpha)^{|\mathcal{N}_i| - \sum a_{\mathcal{N}_i}}$ denote the probability of the nearest neighborhood for an individual i receiving treatment $A_{\mathcal{N}_i}$ under allocation strategy α.
- Define $\bar{y}_i(a,\alpha) = \sum_{a_{\mathcal{N}_i}} y_i(a_i = a, a_{\mathcal{N}_i})\pi(a_{\mathcal{N}_i};\alpha)$ to be the average potential outcome for individual i under allocation strategy α.

Estimands (1)

Under allocation strategy α, the probability of neighborhood of i is denoted by $\pi(a_{\mathcal{N}_i};\alpha) = \alpha^{\sum a_{\mathcal{N}_i}}(1-\alpha)^{d_i - \sum a_{\mathcal{N}_i}}$ and the probability of individual of i is $\pi(a_i;\alpha) = \alpha^{a_i}(1-\alpha)^{1-a_i}$.

The population average potential outcome is defined by

$$\bar{y}(a,\alpha) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a_{\mathcal{N}_i}} y_i(a_i = a, a_{\mathcal{N}_i})\pi(a_{\mathcal{N}_i};\alpha)$$

and the marginal population average potential outcome is

$$\bar{y}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a_i, a_{\mathcal{N}_i}} y_i(a_i, a_{\mathcal{N}_i})\pi(a_i, a_{\mathcal{N}_i};\alpha)$$

Estimands

Under allocation strategy α, the direct effect is

$$\overline{DE}(\alpha) = \bar{y}(1, \alpha) - \bar{y}(0, \alpha).$$

The disseminated or indirect effect under allocation strategy $\alpha = (\alpha_0, \alpha_1)$ is

$$\overline{IE}(\alpha) = \bar{y}(0, \alpha_1) - \bar{y}(0, \alpha_0).$$

The composite or total effect is

$$\overline{TE}(\alpha) = \bar{y}(1, \alpha_1) - \bar{y}(0, \alpha_0).$$

The overall effect is

$$\overline{OE}(\alpha) = \bar{y}(\alpha_1) - \bar{y}(\alpha_0).$$

Assumptions (1)

- The potential outcomes only depends on exposure of the individual and their nearest neighbors $y_i | a_i, a_{\mathcal{N}_i}$.
- Conditional exchangeability for participants:

$$\Pr(A_i = a_i | Z_i = z_i) = \Pr(A_i = a_i | z_i, z_{\mathcal{N}_i}, y_1(\cdot), \ldots, y_n(\cdot))$$

- Conditional exchangeability for neighbors:

$$\Pr(A_i = a_i, A_{\mathcal{N}_i} = a_{\mathcal{N}_i} | z_i, z_{\mathcal{N}_i}) = \Pr(A_i = a_i, A_{\mathcal{N}_i} = a_{\mathcal{N}_i} | z_i, z_{\mathcal{N}_i}, y_1(\cdot), \ldots, y_n(\cdot))$$

Assumptions (2)

- We assume the treatment positivity

$$Pr(a_i, a_{\mathcal{N}_i} | z_i, z_{\mathcal{N}_i}) > 0 \text{ for all } a_i, a_{\mathcal{N}_i}, z_i, \text{ and } z_{\mathcal{N}_i}.$$

- $C_i \perp\!\!\!\perp A_i | L_i$. i.e. C_i only depends on the baseline covariates.
- Stratified interference assumption with nearest neighbors
- Smaller groupings or neighborhoods for each individual can be identified in the observed network.

Nearest Neighborhood IPW Estimator

$$\widehat{Y}^{IPW}(a,\alpha) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i(A_i, A_{\mathcal{N}_i}) I(A_i = a)\pi(A_{\mathcal{N}_i};\alpha)}{f(A_i, A_{\mathcal{N}_i}|Z_i, Z_{\mathcal{N}_i})}.$$

The marginal IPW estimator is defined as

$$\widehat{Y}^{IPW}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i(A_i, A_{\mathcal{N}_i})\pi(A_i, A_{\mathcal{N}_i};\alpha)}{f(A_i, A_{\mathcal{N}_i}|Z_i, Z_{\mathcal{N}_i})}.$$

The propensity score is defined as

$$f(A_i, A_{\mathcal{N}_i}|Z_i, Z_{\mathcal{N}_i}) = \int \prod_{j\in\mathcal{N}_i^*} p_j^{A_j}(1-p_j)^{1-A_j} f(b_i, 0, \theta_s)db_i$$

where $\mathcal{N}_i^* = \mathcal{N}_i \cup \{i\}$ and $p_j = \text{logit}^{-1}(Z_i \cdot \theta_z + b_i)$.

Simulation Results

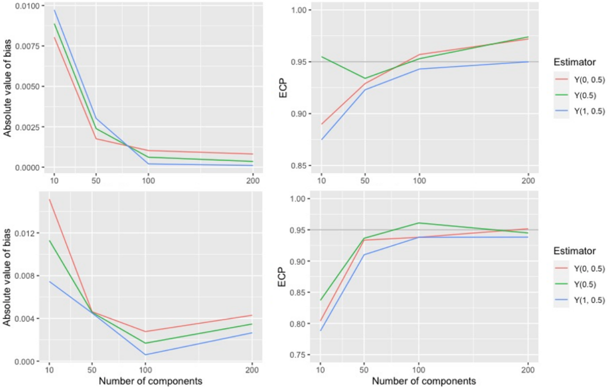


Figure 2: The average absolute value of bias (left) and empirical coverage probability (right) on networks with 10, 50, 100, and 200 components using logistic censoring model (top) and mixed effect censoring model (bottom)

Community Alerts and HIV Risk Behavior in TRIP at 6 months

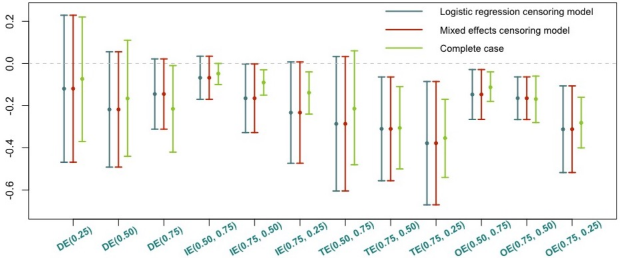


Figure 3: The risk difference estimates and the Wald 95% confidence intervals of direct, indirect, total, and overall effects under allocation strategies 25%, 50%, and 75%.

Acknowledgements

Thank you to my collaborators, Drs. Friedman, Halloran, Katenka and Marshall. The project described was supported by the Avenir Award with Institutional Development Award Number DP2DA046856 from the National Institute on Drug Abuse. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Discussion Questions

- Which settings would you expect there to be interference? Which settings would you find the assumption of no interference plausible?
- Do you have any suggestions on how to disentangle homophily from a causal effect in a network?
- Can you think of any other ways to create (conditional) independence in a network?