ENAR 2020 SPRING MEETING

March 2020

Toward Evaluation of Disseminated Effects of Non-Randomized HIV Prevention Interventions Among Observed Networks of People who Inject Drugs

Ashley Buchanan, Natallia Katenka, TingFang Lee, M. Elizabeth Halloran, Samuel Friedman, and Georgios Nikolopoulos

Introduction

- PWID are embedded in social (HIV/HCV risk) networks and exert biological and social influence on the members of these networks (Hayes et al., 2000; Ghosh et al., 2017).
- In PWID networks, interventions often have **disseminated effects**, which frequently depends on the network structure and intervention coverage levels.

Methodology

- Disseminated effect could be stronger than individual effects and ignoring disseminated effects can under-estimate the full impact of interventions (Buchanan et al., 2018).
- Interference sets could be defined various ways (e.g., study clusters, neighborhoods, nearest neighbors).

Proposed Methodology

Aims to evaluate the disseminated (or indirect, spillover) effect within networks in which PWID are imbedded, while allowing for overlap between communities or neighborhoods in the network.

Motivating Study: Transmission Reduction Intervention Project (TRIP)

- Adult individuals and members of their social networks who are at heightened risk of HIV infection to prevent or treat HIV and other sexually transmitted diseases
- Injection drug users in Odessa, Ukraine and Athens, Greece and men who have sex with men in Chicago, United States from 2013 to 2015
- Treatment as prevention (TasP) and community alerts interventions for newly-infected persons
- Extended contact tracing and two-step trace after a recently infected person

TRIP Community Alerts

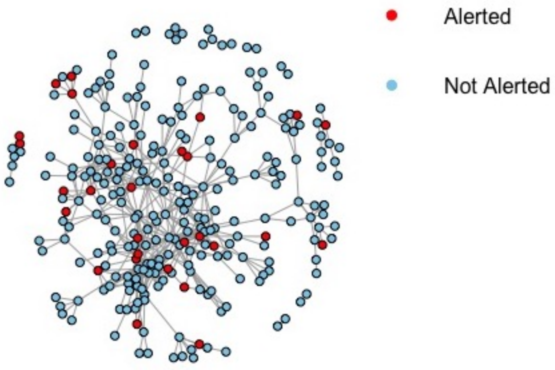

Figure 1: TRIP network with isolates removed (11% exposed from study staff)

Notations and Assumptions

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 exposures for participants in $\mathcal{N}_i$

$Z_{\mathcal{N}_i}$: the vector of covariates for participants in $\mathcal{N}_i$

$y_i(a_i, a_{\mathcal{N}_i})$ denote the potential outcome of participant i if they received treatment a_i and their nearest neighbors received $a_{\mathcal{N}_i}$.
$Y_i(A_i, A_{\mathcal{N}_i})$ be the observed outcome.

Assumptions

- We assume the sufficient conditions for valid estimation of causal effects, i.e. conditional exchangeability, consistency, and positivity.
- Smaller groupings or neighborhoods for each individual can be identified in the data.
- The potential outcomes of a participant only depends on the exposure of his/her nearest neighbors and not the exposures of any other individuals in the network.
- Stratified interference assumption: the potential outcomes of a participant depends on the proportion of the exposure of his/her nearest neighbors and not the vector of exposures of the nearest neighbors.

Estimands

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

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

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 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

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

IPW Estimator I

Given an individual i and $\mathcal{N}_i$, define $\mathcal{N}_i^* = \mathcal{N}_i \cup \{i\}$, $\mathcal{N}_{i,-j}^* = \mathcal{N}_i^* \backslash \{j\}$, and $d_i^* = |\mathcal{N}_i^*| = d_i + 1$. We define the weight that considers the second-order neighbors for each observed outcome

$$w_i = \sum_{j \in \mathcal{N}_i^*} \frac{\pi(A_{\mathcal{N}_{j,-i}^*;\alpha})}{d_j^* f_1(A_j, A_{\mathcal{N}_j} | Z_j, Z_{\mathcal{N}_j})}.$$

Here $f_1(A_i, A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i})$ is the group/neighborhood level propensity score given by

$$f_1(A_i, A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i}) = \int_{-\infty}^{\infty} \prod_{j \in \mathcal{N}_i^*} p_j^{A_j} (1 - p_j)^{1 - A_j} \text{dnorm}(b_i, 0, \psi) db_i$$

where $p_j = \text{logit}^{-1}(Z_j \cdot \gamma + b_i)$

IPW Estimator I

The population-level and marginal population-level average potential outcome estimators are

$$\widehat{Y}^{IPW_1}(a;\alpha) = \frac{1}{n}\sum_{i=1}^{n} Y_i(A_i, A_{\mathcal{N}_i}) I(A_i = a) \cdot w_i$$

and

$$\widehat{Y}^{IPW_1}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} Y_i(A_i, A_{\mathcal{N}_i}) \cdot w_i.$$

Simulation

Potential outcomes

Given a known network, we simulate the potential outcomes for each individual

$$y_i(a_i, a_{\mathcal{N}_i}) \sim \text{Bern}(\text{logit}^{-1}(\beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot \frac{\sum a_{\mathcal{N}_i}}{d_i} + \beta_3 \cdot a_i \cdot \frac{\sum a_{\mathcal{N}_i}}{d_i} + \beta_4 \cdot Z_i))$$

where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (-1.75, 0.5, 1, -1.5, 1)$.
We then get true direct and disseminated effects from the potential outcomes.

Simulation

Observed data set

We first generate the treatment/exposure with a propensity score mixed effect model

$$A_i \sim \text{Bern}(\text{logit}^{-1}(0.4 + 0.2 \cdot Z_i + \text{random effect}))$$

where $Z_i \sim \text{Bern}(0.5)$.
The observed outcome, $Y_i(A_i, A_{\mathcal{N}_i})$, is the set of outcomes in the potential outcome which corresponds to the combination of the treatment that we generated.

Bootstrap strategies

1. Individual based: Sample individual-level average potential outcome estimators with replacement.
2. Neighborhood based: For each sampled individual, we include the first order neighborhood's individual-level average potential outcome estimators.
3. Community based:
   a. Use community detection to define the clusters
   b. Obtain the community-level potential outcome estimators by averaging the individual-level potential outcome estimators within each cluster.
   c. Bootstrap the community-level average potential outcome estimators.

ENAR
ENAR 2020 SPRING MEETING

Simulation Results

Table 1: Simulation results for estimator I of direct, disseminated, composite and overall effects for 1,000 simulated data sets.[a]

Parameter	Bias	ESE	ASE1	ASE2	ASE3	ECP	ECP1
$RD^D(.25)$	0.044	0.104	0.124	0.065	0.178	0.94	0.99
$RD^D(.50)$	0.013	0.067	0.095	0.051	0.135	0.96	0.99
$RD^D(.75)$	-0.023	0.072	0.100	0.053	0.144	0.95	0.99
$RD^I(.25, .50)$	-0.020	0.061	0.047	0.028	0.078	0.96	0.95
$RD^I(.50, .75)$	-0.025	0.035	0.044	0.025	0.071	0.89	0.94
$RD^I(.25, .75)$	-0.045	0.081	0.083	0.048	0.142	0.94	0.97
$RD^T(.25, .50)$	0.025	0.081	0.113	0.056	0.146	0.95	0.99
$RD^T(.50, .75)$	-0.012	0.072	0.101	0.050	0.132	0.96	0.99
$RD^T(.25, .75)$	-0.001	0.086	0.118	0.058	0.149	0.96	0.99
$RD^O(.25, .50)$	-0.015	0.051	0.044	0.027	0.022	0.96	0.95
$RD^O(.50, .75)$	-0.001	0.024	0.032	0.022	0.038	0.96	0.99
$RD^O(.25, .75)$	-0.016	0.066	0.069	0.045	0.055	0.95	0.98

[a] ESE = empirical standard error; ASE = average estimated standard error (individual, neighborhood, community bootstrap); ECP = empirical coverage probability).

Community Alerts and HIV Risk Behavior in TRIP

Effect	Coverage (α, α')	Unadjusted RD	Unadjusted 95% CI	Adjusted[a] RD	Adjusted[a] 95% CI
Direct	(25%, 25%)	0.08	(-0.28,0.44)	-0.09	(-0.40, 0.21)
Direct	(50%, 50%)	-0.04	(-0.6,0.51)	-0.15	(-0.49, 0.19)
Direct	(75%, 75%)	-0.17	(-0.58,0.24)	-0.16	(-0.36, 0.04)
Indirect	(75%, 25%)	-0.05	(-0.13,0.03)	-0.16	(-0.26, -0.06)
Indirect	(50%, 25%)	0.004	(-0.06,0.07)	-0.06	(-0.12, 0.001)
Indirect	(75%, 50%)	-0.06	(-0.11,-0.003)	-0.10	(-0.15, -0.04)
Total	(75%, 25%)	-0.22	(-0.58,0.13)	-0.31	(-0.59, -0.04)
Total	(50%, 25%)	-0.04	(-0.41,0.33)	-0.21	(-0.50, 0.08)
Total	(75%, 50%)	-0.23	(-0.77,0.31)	-0.25	(-0.58, 0.07)
Overall	(75%, 25%)	-0.20	(-0.28,-0.12)	-0.25	(-0.34, -0.17)
Overall	(50%, 25%)	-0.04	(-0.12,0.04)	-0.11	(-0.16, -0.06)
Overall	(75%, 50%)	-0.16	(-0.22,-0.1)	-0.14	(-0.20, -0.09)

[a] Model adjusted for HIV test, HIV status, age, education, employment, shared drug equipment in the last 6 months.

Next Steps

- Continue to work on asymptotic variance derivations and simulation study
- How to handle varying neighborhood sizes (Basse, 2018)
- Doubly-robust approaches
- Missing outcome information at six-month visit

Selected Reference

- Eric J Tchetgen Tchetgen and Tyler J Vander Wheele, On Causal inference in the presence of interference, *Statistical Methods in Medical Research*, 21(1):55-75, 2012
- Ashley L Buchanan, Sten H Vermund, Samuel R Friedman, and Donna Spiegelman, Assessing individual and desseminated effects in network-randomized studied, *2018*
- Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli, Identification and estimation of treatment and interference effects in observational studies on networks, *arXiv preprint arXiv:1609.06245, 2016.*
- Bradley C. Saul, Michael G. Hudgens, A Recipe for inferference: Start with Causal Inference. Add Interference. Mix Well with R, *Journal of Statistical Software, Articles, v82i02, 2017*

Acknowledgements

The project described was supported by the Avenir Award with Institutional Development Award Number 1DP2DA046856-01 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.

Random effects simulation

We first generate individual random effects $b_i \sim N(0, 0.5)$. The neighborhood level random effect

$$b_i^* = \frac{\sum_{j \in \mathcal{N}_i^*} b_j}{\sqrt{d_i^*}}$$

of each individual is assigned as the average of the neighborhood's individual random effects.

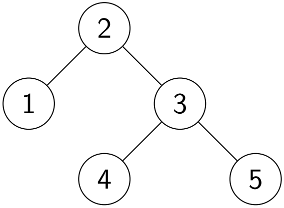

Figure 2: Sample network

$$b_1^* = (b_1 + b_2)/\sqrt{2}$$
$$b_2^* = (b_1 + b_2 + b_3)/\sqrt{3}$$
$$b_3^* = (b_2 + b_3 + b_4 + b_5)/\sqrt{4}$$
$$b_4^* = (b_3 + b_4)/\sqrt{2}$$
$$b_5^* = (b_3 + b_5)/\sqrt{2}$$

Bootstrap estimators

Each simulated observed data, we obtained the individual-level and marginal population-level average potential outcome estimators

$$\widehat{Y}_i^{IPW_1}(a;\alpha) = y_i(A_i, A_{\mathcal{N}_i})I(A_i = a)w_i,$$

$$\widehat{Y}_i^{IPW_1}(\alpha) = y_i(A_i, A_{\mathcal{N}_i})w_i,$$

$$\widehat{Y}_i^{IPW_2}(a;\alpha) = \frac{y_i(A_i, A_{\mathcal{N}_i}I(A_i = a)\pi(A_{\mathcal{N}_i;\alpha})}{f_2(A_i, \sum A_{\mathcal{N}_i}|Z_i, Z_{\mathcal{N}_i})},$$

and

$$\widehat{Y}_i^{IPW_2}(\alpha) = \frac{y_i(A_i, A_{\mathcal{N}_i})\pi(A_i, A_{\mathcal{N}_i;\alpha})}{f_2(A_i, \sum A_{\mathcal{N}_i}|Z_i, Z_{\mathcal{N}_i})}.$$

Exposure and outcome in TRIP study

Exposure

- **Community Alerts from Study Staff**

$$A_{ij} = \begin{cases} 1, & \text{Received} \\ 0, & \text{Did not receive} \end{cases}$$

Outcome

- **Report of Injection HIV Risk Behavior at 6-month Visit**

$$Y_{ij} = \begin{cases} 1, & \text{At least some} \\ 0, & \text{None reported} \end{cases}$$

Baseline covariates in TRIP study

- Have been tested for HIV before
- HIV test status
- Age
- Education status
- Employment status
- Shared drug equipment in the past 6 months

IPW Estimator II

Define the inverse probability weighted estimator for treatment a with coverage α of an intervention the neighborhood as

$$\widehat{Y}^{IPW_2}(a,\alpha) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i(A_i, A_{\mathcal{N}_i}) I(A_i = a)\pi(A_{\mathcal{N}_i};\alpha)}{f_2(A_i, \sum A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i})}$$

and the inverse probability weighted marginal estimator as

$$\widehat{Y}^{IPW_2}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i(A_i, A_{\mathcal{N}_i})\pi(A_{\mathcal{N}_i};\alpha)}{f_2(A_i, \sum A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i})}.$$

IPW Estimator II

The propensity score $f_2(A_i, \sum A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i})$ can be written as

$$f_2(A_i, \sum A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i}) = f_2(\sum A_{\mathcal{N}_i} | A_i, Z_{\mathcal{N}_i}) f_2(A_i | Z_i)$$
$$= \binom{n_i}{\sum A_{\mathcal{N}_i}} p_{1,i}^{\sum A_{\mathcal{N}_i}} (1 - p_{1,i})^{n_i - \sum A_{\mathcal{N}_i}} \cdot p_{2,i}^{A_i} (1 - p_{2,i})^{1 - A_i}$$

where $p_{1,i} = \text{logit}^{-1}(\frac{\sum Z_{\mathcal{N}_i}}{n_i} \cdot \gamma_1)$ and $p_{2,i} = \text{logit}^{-1}(Z_i \cdot \gamma_2)$.

Simulation Results for Estimator II

Table 2: Simulation results for estimator II of direct, disseminated, composite and overall effects for 1,000 simulated data sets.[a]

Parameter	Bias	ESE	ASE1	ASE2	ASE3	ECP	ECP1
$RD^D(.25)$	-0.008	0.138	0.123	0.052	0.184	0.95	0.99
$RD^D(.50)$	-0.025	0.047	0.06	0.02	0.115	0.94	0.98
$RD^D(.75)$	-0.036	0.068	0.084	0.034	0.14	0.92	0.94
$RD^I(.25, .50)$	-0.043	0.068	0.049	0.021	0.075	0.94	0.96
$RD^I(.50, .75)$	0.017	0.048	0.052	0.024	0.074	0.94	0.98
$RD^I(.25, .75)$	-0.025	0.099	0.094	0.042	0.144	0.95	0.98
$RD^T(.25, .50)$	-0.051	0.113	0.102	0.04	0.155	0.95	0.99
$RD^T(.50, .75)$	-0.008	0.067	0.084	0.032	0.134	0.96	0.97
$RD^T(.25, .75)$	-0.033	0.124	0.121	0.05	0.171	0.96	0.99
$RD^O(.25, .50)$	-0.032	0.058	0.044	0.02	0.068	0.94	0.97
$RD^O(.50, .75)$	0.032	0.025	0.029	0.013	0.052	0.77	0.86
$RD^O(.25, .75)$	1e-4	0.073	0.067	0.031	0.111	0.96	0.95

[a] ESE = empirical standard error; ASE = average estimated standard error (individual, neighborhood, community bootstrap); ECP = empirical coverage probability).