2018

EVALUATING THE EFFECTS OF ATTITUDES ON HEALTH-SEEKING BEHAVIOR AMONG A NETWORK OF PEOPLE WHO INJECT DRUGS

Ayako Miura
University of Rhode Island, ayakoshmd@gmail.com

EVALUATING THE EFFECTS OF ATTITUDES ON HEALTH-SEEKING BEHAVIOR AMONG A NETWORK OF PEOPLE WHO INJECT DRUGS

BY

AYAKO MIURA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

STATISTICS

UNIVERSITY OF RHODE ISLAND

2018

MASTER OF SCIENCE IN STATISTICS THESIS

OF

AYAKO MIURA

APPROVED:

Thesis Committee:

Major Professor Natallia Katenka

Ashley Buchanan

Prabhani Kuruppumullage Don

Stephen Kogut

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2018

ABSTRACT

The transmission of HIV/AIDS remains a great concern among people who inject drugs (PWIDs) in the United States. PWIDs are often embedded in a unique HIV/AIDS risk network via the shared use of drug equipment and risky sexual behavior. However, the characteristics of PWIDs in risk networks present challenges in a collection of network data resulting in limited studies of these networks of PWIDs. Our study employed causal inference methods applied to an observational study with dissemination to assess attitudes toward HIV/AIDS risk among PWIDs and their effect on health-seeking behaviors.

We used data from the Social Factors and HIV Risk Study (SFHR), a sociometric network study conducted between 1991 and 1993 in Bushwick, Brooklyn, New York that investigated how HIV/AIDS infection spread among PWIDs through shared sexual and injection risk behaviors. We evaluated the effects of locus of control (internal vs. external) and blame (self vs. others) attitudes separately on their own health-seeking behavior and that of other members in their risk communities. With taking dissemination of attitudes into account, four causal parameters were estimated: *direct*, *indirect*, *total*, and *overall* effects. Communities were defined to include members that were closely related via HIV risk behavior and had sparser connections with individuals outside of the community. For the health-seeking behavior outcomes, we considered receipt of study-based HIV testing result and a medical encounter within the past year. While *direct* effect measures direct effect of exposure on outcome behavior of PWIDs in the same community, *indirect* effect is the quantified measure of dissemination, which compares the outcomes of unexposed PWIDs in two

different communities. *Total* effect is defined as the sum of *direct* and *indirect* effects and is the measure of the maximum impact of the exposure of interest. Last but not least, *overall* effect measures the marginal effect of exposure by comparing the potential outcomes of those exposed and unexposed regardless of communities they belong.

First, we applied a modularity-based community detection algorithm to determine communities within the SFHR network. We then employed a network-based causal inference methodology for clustered observational data. Coverage is defined as the proportion of people with internal locus of control/self-blame attribute in a community. For the *direct* effect, PWIDs who believe uncontrollable factors determine whether or not one gets HIV/AIDS (i.e. with external locus) were 16% less likely to receive HIV testing result when they are in 50% and 70% coverage communities (95% confidence interval (CI): -0.27, -0.06, for both communities). Also, when the coverage of people who believe controllable factors determine whether or not one gets HIV/AIDS (i.e. with internal locus) was decreased from 70% to 50%, the likelihood of receiving HIV testing result decreases 3% among those with external locus (95% CI: -0.05, -0.01), demonstrating a significant disseminated effect. Furthermore, as another significant dissemination effect, when the coverage of people with self-blame was decreased from 99% to 50%, the likelihood of having a recent medical encounter increases 27% for those with external locus (95% CI: 0.07, 0.47).

Because the SFHR study was conducted in the early 1990s, and there is a possibility that the health-seeking behavior of PWIDs has somewhat changed over time. However, our results may contribute to understanding how PWIDs attitudes and

behaviors have changed over a few decades by conducting a similar analysis in more contemporary studies. The results from this study support the existence of dissemination of locus of control/blame attitudes among PWID networks. This indicates that the introduction of appropriate network-targeted interventions can bolster positive behavioral change in health-seeking among PWIDs by leveraging disseminated effects.

ACKNOWLEDGMENTS

I would first like to express my profound gratitude to my advisors Dr. Natallia V. Katenka and Dr. Ashley L. Buchanan for their patience, encouragement and thoughtful directions during my research and writing.

Besides my advisors, I would like to thank the rest of my thesis committee and the chair: Dr. Prabhani Kuruppumullage Don, Dr. Stephen Kogut, and Dr. Xuerong Wen, for their encouragement, insightful comments. I would also like to thank Dr. Samuel R. Friedman at National Development and Research Institutes, Inc.

for allowing me to use the SFHR study, which are supported by NIH grant R01DA006723, and providing me very valuable comments on my research.

My sincere thanks also go to Dr. Liliana Gonzalez, Dr. Gavino Puggioni, Dr. Steffen Ventz, and Dr. Jing Wu for providing me helpful advice and support to my study and future career. I thank my friends: Hilary Aroke, Katherine Abramski, Kaitlin Dio and Caoxin Sun for discussing topics in Statistics and research.

Finally, I would like to thank my husband for providing continuous encouragement throughout my years of study and through the process of researching and writing this thesis. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The transmission of HIV/AIDS remains a great concern among people who inject drugs (PWIDs) in the United States. Although the Centers for Disease Control and Prevention (CDC) reports that the number of people who are diagnosed as HIV has declined during the last decade overall, certain groups continue to experience a disproportionate burden of HIV. Ascertainment of the population of PWIDs is often difficult because of illegitimacy and stigma (1). Because injection drug use increases the risk of HIV transmission via sharing injection drug equipment and often correlates with engagement in risky sexual behaviors, there are continued efforts to improve HIV treatment and prevention among PWIDs. Socioeconomic factors often experienced by PWIDs can increase barriers to accessing HIV prevention and treatment. For example, in cities with high HIV prevalence, more than half of HIV-positive PWIDs were homeless, 30% were incarcerated and 20% had no health insurance in the last 12 months[1]. PWIDs are also a key population in terms of global HIV/AIDS prevention. Ghosh et al. reports that PWIDs are a key population facing unique barriers along the HIV continuum care due to the absence of supporting programs with firm financial background (2). Thus, PWIDs are not only at a higher risk of HIV infection, but also face unique barriers on the continuum of HIV care.

Among PWIDs, injection drug use can be a pathway of HIV transmission via shared injection drug use equipment and/or engagement in risky sexual behavior.

[1] https://www.cdc.gov/hiv/group/hiv-idu.html (accessed on June 19, 2018)

Facing these multiple risks factors, standard HIV prevention efforts may not be as effective possibly due to the difficulty to completely identify this target subpopulation. Furthermore, socioeconomic factors among PWIDs, such as poverty and lack of access to sufficient medical care, often prevent individuals from engaging in the subsequent steps along the HIV continuum of care, which include diagnosis of HIV, linkage to care and retention among HIV-positive individuals, adhering to antiretroviral therapy (ART) and, among those on ART, well-controlled viremia (2).

Network-targeted interventions that can be strengthened by community resonance (or dissemination) offer a possibility of more effective and sustainable solutions for HIV/AIDS prevention among PWIDs. Dissemination is often discussed in the context of vaccination to infectious disease with herd immunity, which is the concept that individuals who are vaccinated protect not only themselves but also other individuals in the same community from acquiring the disease.

Similar benefits in network-targeted interventions where people are connected by having a direct contact and/or sharing information are anticipated. Curtis *et al*. investigated the characteristics of PWIDs network and HIV risk in a network ascertained from Bushwick, Brooklyn, New York in the early 1990s (3) and defined three different categories of injection drug users: core network members (i.e., people who have direct connections with drug distribution market), inner periphery (i.e., people who have direct connections with core members by sharing drugs within 30 days before the interview) and outer periphery (i.e., people who do not depend on core members to obtain information of drugs and place where they shoot drugs). Among these participants, there was a significant difference in risk behavior between outer

periphery and the other two categories, and members of the outer periphery used injection drugs and engaged in risky behaviors less frequently compare to other two types of injection drug users (3). HIV/AIDS prevention that appreciates the network structure in which PWIDs are imbedded, and particularly those that target core network members and those on the inner periphery could be more effective compared to the conventional approach, which typically encourages all network members to modify their risk behavior related to injection drug user.

The network structure in which PWIDs are embedded could slow or even prevent them from attaining long-term behavioral change, or vice versa, improve and sustain behavioral change (3). Introducing appropriate network-targeted interventions can bolster positive behavioral change among PWIDs (4). The purpose of this study is to investigate how personal health attitudes impact health-seeking behavior among PWIDs and their HIV risk networks. This new information can then be used to develop more effective interventions for HIV/AIDS treatment and prevention among injection drug users, such as an educational program to empower individuals to engage in health behavior targeted at the most influential members in communities of PWIDs.

One challenge to evaluate the effect of attitudes of PWIDs on their health-seeking behaviors is dissemination or interference among the PWIDs network. That is, the health beliefs and/or blame attributes of an individual can affect the health-related behaviors of their network members. Using the potential outcome framework for causal inference, one typically assumes no dissemination of the treatment or exposure; that is, an individual's outcome is affected only by their own treatment/exposure and not by the treatment/exposure received by other individuals in the study. This

assumption is part of the "stable unit-treatment value assumption" (SUTVA) (5). However, in some settings, dissemination is of interest to understand causal relationships and the full impact of an intervention. For example, consideration of interference is necessary to evaluate the effect of vaccination for the prevention of infectious disease (6, 7). In our setting where we try to evaluate the effect of PWIDs attitudes toward HIV/AIDS risk, ones' attitude can change another person's health-seeking behavior especially when they are closely connected. For the estimation of causal inference in the presence of interference, we use an inverse probability weighting (IPW) method developed by Tchetgen Tchetgen and VanderWele (8) designed to replicate an idealized two-stage randomized design. In this design, investigators randomly assign a treatment allocation strategy to each community then assign actual treatment to participants in each community given the assigned treatment allocation strategy. When applying this method to observational studies, group-level propensity scores are computed for each observed community in population of interest. Lastly, the inverse of this propensity score is used as weight for the IPW estimator that is a contrast of group-level potential outcomes.

An individual's health attitudes can be defined based on two distinct concepts of locus of control and blame attribute. The concept of locus of control was developed by Rotter in the field of personal psychology (9). The locus of control is defined as the degree of people's belief that how much they have control on what happen to them. Locus of control is classified into two different types: internal and external (9). People with internal locus attribute the events they experience to factors within their control while those with external locus attribute events to factors beyond their own influence.

The rational for a distinction between locus of control and blame attribute has precedent in the field of psychology. Blanchard-Fields *et al*. (2012) investigated how one's beliefs against traditional social schema affects his/her blame attribute against the violation of such schema by others (10). Grimes *et al*. (2004) investigated how students' locus of control affect the evaluation of their teacher (11). They found that students who had internal control tended to give high evaluation to their teacher while students with external locus gave lower evaluation to the teachers. Locus of control and blame attribute have been considered distinct concepts in psychology. Among PWIDs, there have been few studies on the relationships between locus of control and blame attribute with health-seeking behavior. We considered locus of control and blame attribute as distinct exposures and the evaluated the exposure effects on health-seeking behavior among each individual and their networks.

List of References

1. Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, et al. Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. *The Lancet*. 2008;372(9651):1733-45.

2. Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social Network Strategies to Address HIV Prevention and Treatment Continuum of Care Among At-risk and HIV-infected Substance Users: A Systematic Scoping Review. *AIDS and Behavior*. 2017;21(4):1183-207.

3. Curtis R, Friedman SR, Neaigus A, Jose B, Goldstein M, Ildefonso G. Street-level drug markets: Network structure and HIV risk. *Social Networks*. 1995;17(3-4):229-49.

4. Valente TW. Network interventions. *Science*. 2012 Jul 6;337(6090):49-53.

5. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Review of Public Health*. 2000;21(1):121-45.

6. Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. *Biometrics*. 2014;70(3):731-41.

7. Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology*. 1991:331-8.

8. Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. *Statistical Methods in Medical Research*. 2012;21(1):55-75.

9. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*. 1966;80(1):1.

10. Blanchard-Fields F, Hertzog C, Horhota M. Violate my beliefs? Then you're to blame! Belief content as an explanation for causal attribution biases. *Psychology and Aging*. 2012;27(2):324.

11. Grimes PW, Millea MJ, Woodruff TW. Grades—Who's to blame? Student evaluation of teaching and locus of control. *The Journal of Economic Education*. 2004;35(2):129-47.

CHAPTER 2

METHODS

2.1 Study data

In this study, we evaluated how PWIDs' individual health beliefs affect their own health-seeking behavior and their risk network members' health-seeking behavior in the Social Factors and HIV Risk Study (SFHR). The SFHR study was conducted in Brooklyn and other parts of New York, New Jersey and Connecticut between July 1991 and January 1993 (1). Data was collected from street recruited injection drug users in the Bushwick neighborhood, a low-income area of approximately 100,000 residents with high rates of poverty, injection drug use, and HIV/STI prevalence. The original study enrolled a total of 767 participants and the information of 3,162 dyadic relationships, a connection between two individuals. HIV risk connections were defined by sharing risk behaviors (i.e. use drug together or have sexual intercourse).

2.2 Data preparation

Our primary objective was to understand how injection drug user's individual attitudes affect health-seeking behaviors of that individual and those with shared HIV risk connections in the network structure. From this perspective, our analysis focused on the individuals who had at least one shared risk connection with participants in SFHR study. We also define the participants who do not share any risk links with at least one other enrolled participants as isolated participants. The process of sample

selection for the analysis in our study was summarized in a flowchart (Figure 1). After removing 82 participants either missing outcome, exposure or covariate information and 283 isolated participants, the SFHR PWIDs network for this analysis included 402 subjects with 403 risk connections (Figure 2).

Figure 1. Flowchart of Study Sample Selection.

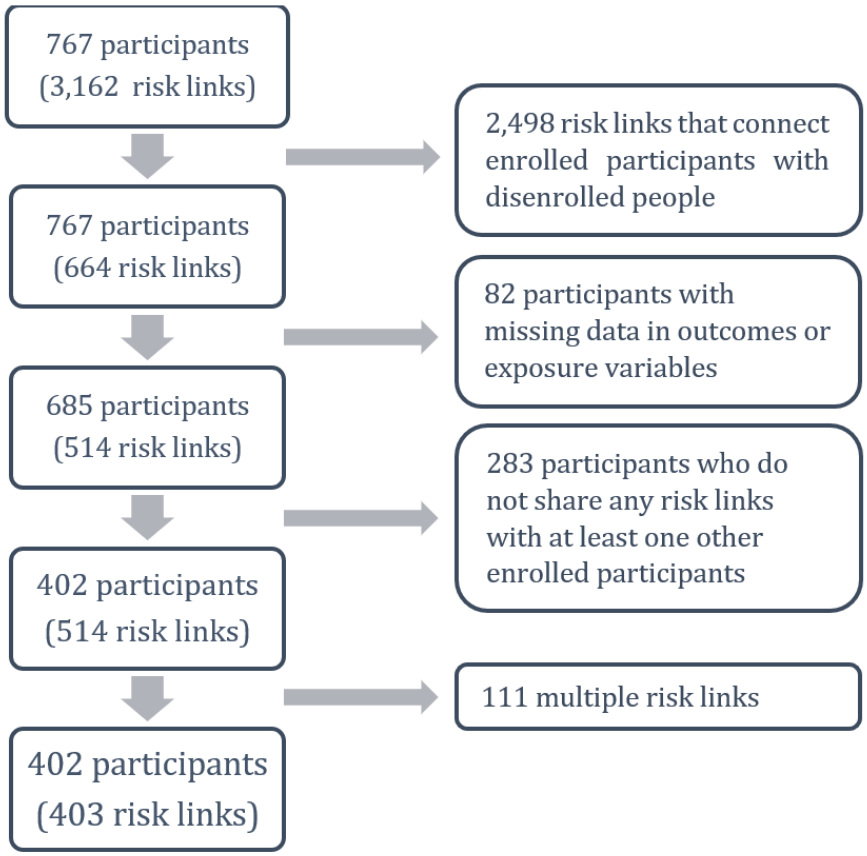


Figure 2. The Social Factors and HIV Risk Study People Who Inject Drugs' Network for the Analysis with 402 Nodes and 403 Links. The Colors of Nodes Represent Locus of Control/Blame Attributes and Shapes of Nodes Show Individual Health Seeking Behavior.

To evaluate the impact of health beliefs, we defined the individual's health belief status based on available information in SFHR. In our context, we classified PWID's individual health beliefs by locus of control with respect to HIV/AIDS as following: an individual with *internal locus* believes that controllable factors, such as

own effort or action, determine whether one gets HIV/AIDS and an individual with *external locus* believes that uncontrollable factors, such as bad luck or fate, determine whether one gets HIV/AIDS. In SFHR study, participants were asked questions to measure their health beliefs with respect to HIV/AIDS. Table 1 lists all ten statements asked regarding health beliefs and participants responded how much they agreed with each statement (i.e. strongly agree, agree, somewhat agree, disagree, strongly disagree, and don't know)[2]. While some questions (i.e. *Question 3. I'm in control of whether or not I get AIDS.*) ask about locus of control, other questions (*Question 9. If I get AIDS it is because of the society we live in*.) focus on the issue of blame for their HIV/AIDS status. Questions 4, 5 and 9 capture information about PWIDs' individual blame (self-blame or blame-others) and the remaining questions are about PWIDs' HIV/AIDS locus of control (internal or external locus).

The original responses to belief and blame questions in SFHR study were recorded on a Likert scale. To create a binary variable to represent individual's health belief status based on these categorical responses, for each response out of seven heal belief related questions, we assigned value of 1 if a person showed internal locus of control in its response while values of -1 was assigned if the subject's response showed external locus of control. If the response was neutral (i.e. don't know), we assigned 0. This rescaling was conducted because there is a problem in assign integer values to the original response categories asked in SFHR. The problem is that the original response categories are not symmetric or balanced. That is, there is not somewhat disagree in the categories. As mentioned in Andrich (1978), integer-scoring

[2] The original response to the SFHR health belief question included "refused" and "not applicable" as option. However, not these answers were observed in the participants' responses.

approach requires the assumption of equal distances between categories (2). Therefore, by rescaling with -1, 0 and 1, we could satisfy this assumption and create a score for each participant. By adding the all values assigned to the responses for health belief related question, we obtained individual health belief score (BLF score) ranging from -7 to 7. The distribution of BLF scores is shown in the left panel of Figure 3. This integer-scoring procedure is common approach in psychological studies in dealing with categorical responses (2, 3). Then, if one's BLF score was greater than or equal to three, we considered the overall health belief status of that person is internal locus of control otherwise the person's health belief status is external locus (i.e. $a_{BLF} = 1$ for internal locus; $a_{BLF} = 0$ for external locus). A similar procedure was implemented to create a binary variable to represent individual blame attribute. By summation, we obtained blame scores that ranged from -3 to 3 for individual participant. The distribution of BLM scores is shown in the right panel of Figure 2. If one's blame score was equal to three, the attribute was "self-blame"; otherwise, the attribute was "blame others" (i.e. $a_{BLM} = 1$ for self-blame; $a_{BLM} = 0$ for blame others). We used a threshold of three for defining individual locus of control as a binary variable and we chose this threshold because loglikelihood was higher in case of using three compares to use another threshold such as five or seven (The loglikelihoods from the estimation of propensity score with the threshold of 3, 5 and 7: -171; -242; -261, respectively).

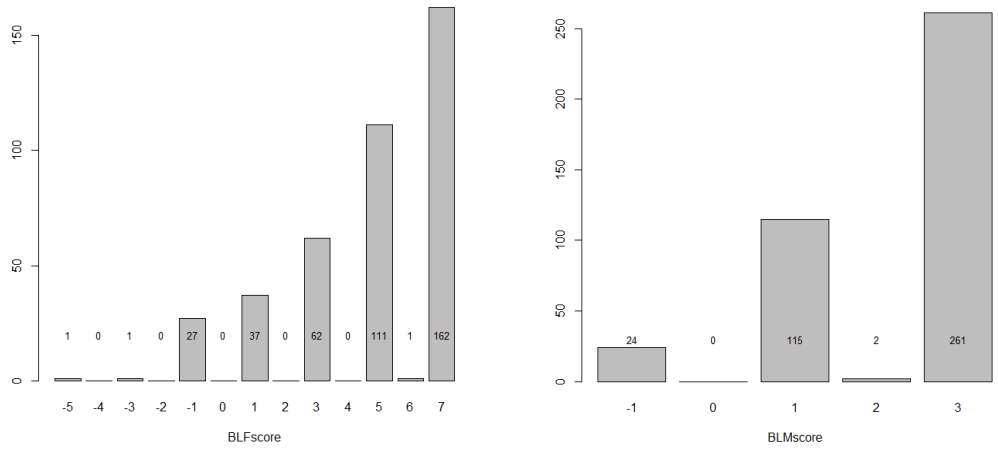
We evaluated two outcomes related to PWIDs' health-seeking behaviors: receipt of SFHR HIV test result and a medical visit within the past year. The HIV test was conducted as a part of interview. Receipt of HIV test for individual participant

was a binary variable (i.e. $Y_P = 1$ if "Yes" or $Y_P = 0$ if "No"). The medical visit status was self-reported and considered as a dichotomous variable (i.e. $Y_D = 1$ if "Yes" or $Y_D = 0$ if "No").

Table1. Questions About Health Beliefs in the Social Factors and HIV Risk Study from Bushwick, New York (1991-1993).

SFHR Questions about Health Beliefs	
Q1. It is my own behavior which determines whether I get AIDS or not.	
Q2. No matter what I do, if I'm going to get AIDS, I will get AIDS.	
Q3. I'm in control of whether or not I get AIDS.	
Q4. My family has a lot to do with whether I get AIDS.	**(Blame)**
Q5. If I get AIDS, I'm to blame.	**(Blame)**
Q6. Getting AIDS is largely a matter of bad luck.	
Q7. No matter what I do, I'm likely to get AIDS.	
Q8. If I take the right actions, I can avoid getting AIDS.	
Q9. If I get AIDS it is because of the society we live in.	**(Blame)**
Q10. No matter what I do, I'm unlikely to get AIDS.	

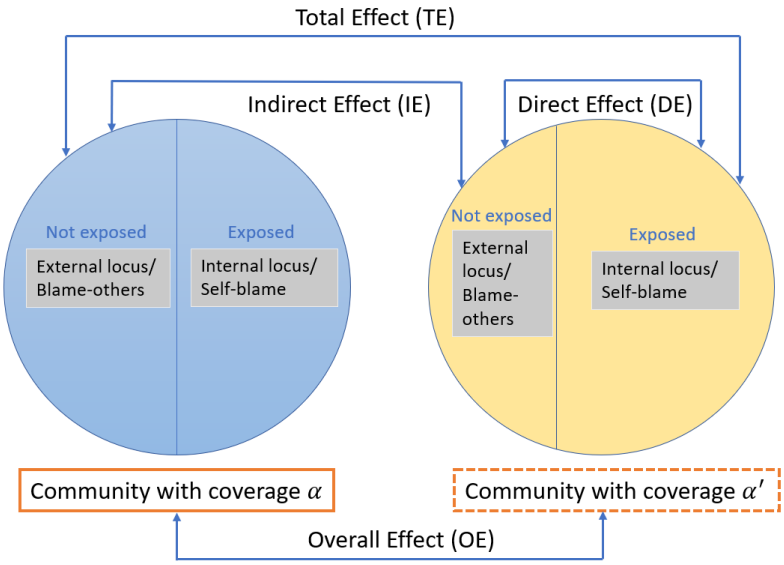
Figure 3. Frequency Distribution of BLF/BLM Scores of Individual Participant Created Based on Health Belief Questions Asked in the Social Factors and HIV Risk Studies.



2.3 Causal inference framework under the presence of interference

Figure 4 is a schematic diagram of causal parameters when dissemination exists in a study and is based on the diagram introduced by Halloran and Struchiner (4). Coverage is the proportion of people who received treatment/exposure in a group. In two-stage randomized design, first, different coverages of exposure (α, α' where $\alpha < \alpha'$) are randomly assigned to different groups or communities, then people in each group are assigned to the exposure according to the pre-assigned coverage strategy to the group. As the result, different coverage groups have both exposed and non-exposed member within a group. Given this particular design, Halloran and Struchiner defined the four different causal effects: Direct effects (DE), Indirect causal effects (or Disseminated effects) (IE), Total causal effects (TE) and Overall causal effects (OE) (4, 5). The *direct effect*, or individual effect, compares potential outcomes of those with and without exposure within a certain coverage group. The *indirect effect*, or disseminated effect, compares potential outcomes of those without exposure who belong to different coverage groups. The *total effect*, or composite effect, is the sum of direct and indirect effect and is considered as the maximum impact of an exposure on an individual who received exposure under a certain coverage compared to another coverage group. Finally, the *overall effect* is defined as population-level marginal effect of a coverage compare to another (Figure 4). In the next subsection, we introduce assumptions and notations required for estimating these causal effects with observational studies.

Figure 4. Schematic Diagram of Causal Estimation when Dissemination Exists. Based on the Diagram Introduced by Halloran and Struchiner (1991).



2.4 Assumptions and notation

An important feature of our study is that we need to allow for the exposure (i.e. individuals' attitudes) to have influence on the behavioral outcomes of other individuals in the same community; however, we assumed this effect does not extend beyond that particular community. This assumption is known as *partial interference* (5, 6, 7). Under this condition, the no interference assumption of SUTVA is relaxed within clusters. We also assume the following three assumptions required to guarantee the internal validity of causal estimation in observational data: i) Conditioning on a set of pre-treatment covariates assumed to be sufficient to control for confounding; that is, the potential outcomes of those who were exposed and the outcomes of those who not exposed are the same on average (*conditional exchangeability*), ii) there is a positive probability of exposure within each level of the covariates (*positivity*), and iii) the exposure is well defined and there is no other version of exposure in the study

(*consistency*). To define the potential outcomes, we assumed a Bernoulli individual group assignment strategy under which each individual has exposed (i.e. having internal locus of control/self-blame attribute) at random with probability α (6). In this study, we also assumed that there is no misclassification of attitudes; that is, every participant tells the truth about his/her attitudes in the study. With relating this, we will make a detailed discussion on reliability and validity of health beliefs questions in SFHR and our definition of exposures in discussion section. In addition to this, we assumed that the exposure status we defined capture the underlying traits of internal vs. external locus of control and self-blame vs. blame others. We also assume the weight models are correctly specified (e.g., correct functional forms of covariates). We also assumed there is no homophily, which means negating the existence of latent variables with which an individual has a tie with another individual who has the similar characteristics. As discussed in McPherson *et al*. (2001), similar individuals are more likely to be connected in the first place, rather than the intervention or exposure influencing connected individuals (8). Our no homophily assumption presumes that the individual covariates we controlled in our study; sex, race, education level, age and attitudes toward HIV/AIDS risk; locus of control and blame attributes are enough to explain the existence of a tie between a pair of PWIDs and there is no other unobserved characteristic as the source of homophily.

Notation used for describing network structure of the PWIDs HIV/AIDS risk network follows the notation used by Newman (2010) and Kolaczyk *et al.* (2014)(9, 10). Notation used for explaining causal inference under the presence of interference follows the notation used by Saul *et al.* (2017) and Tchetgen Tchetgen and

VanderWeele (2012) (5, 11). A network (or graph) is mathematically defines as a collection of vertices (V) and edges (E), $G = (V; E)$. So, in our context, G is the SFHR PWIDs network in which a vertex represents each subject and an edge represents a shared risk behavior between a pair of subjects. There are K clusters and each of the clusters has n_i individuals for $i = 1, 2, \dots, K$, denoted as $C_i, i = 1, \cdots, K$. then Y_{ij}, A_{ij}, x_{ij} represent observed outcome, actual exposure status and baseline covariate vector of jth individual in cluster i. Also, $\boldsymbol{A_i}$ and $\boldsymbol{X_i}$ are a vector of exposure allocations and matrix of baseline covariates for members within cluster i. Let $\mathcal{A}(n)$ be the set of vectors of all possible exposure allocations of length n. For example, under binary exposure: $a \in \{a = 0, a = 1\}$ where $a = 0$ is control and $a = 1$ is exposure, if a cluster consists of only two individuals, then all possible exposure allocations are represented as $\mathcal{A}(2) \equiv \{(0,0), (0,1), (1,0), (1,1)\}$, that is, there are $2^2 = 4$ patterns. Similarly, if a cluster has n_i individuals the possible exposure allocations can be written as $\mathcal{A}(n_i) \equiv \{(0,0, \dots, 0), (0,0, \dots, 1) \cdots (1,1, \dots, 1)\}$, where the length of each vector is n_i and there are 2^{n_i} patterns of exposure allocations. Furthermore, when there is a restriction in the number of allocations of exposure (i.e. $a = 1$) within cluster i and the number exposed is k (or k individuals out of n_i receive active treatment), the possible exposure allocations are denoted as $\mathcal{A}(n_i, k) \equiv \{(0, \dots, 1, \dots, 1), (0, \dots, 1, \dots 1, 0) \cdots (1, \dots, 1, 0, \dots, 0)\}$, where the length of each vector is n_i and there are $n_i!/(n_i - k)!$ patterns. Also, every element of $\mathcal{A}(n_i, k)$ satisfies $I_{n_i}^T \mathrm{a} = k$. This means there are k people who are exposed. Based on this, we can write the potential outcome for individual j in cluster i as $Y_{ij}\left(A_{ij}, \mathrm{a}_{i,-j}\right) = Y_{ij}(\mathrm{a}_i)$ if the cluster was exposed to $\mathrm{a}_i \in \mathcal{A}$. Here, the individual potential outcome not only

depends on person j's exposure, but also the vector of exposures received by everyone else in cluster in cluster i.

2.5 Community detection

A community (or cluster) is defined as a group of vertices densely connected, with only sparser connection to other groups of vertices (10). Hierarchical clustering is one of the most common methods for community detection (9, 10). In this method, the closest or most similar vertices are combined to form communities with a measure of similarity or connection strength between vertices based on the network structure (10). The most popular one in such measures is modularity (9).

Let G be an observed network and assume there are $\boldsymbol{C} = \{C_1, \dots, C_K\}$ candidate K clusters within the network. We also define $f_{ij} = f_{ij}(\boldsymbol{C})$ as the fraction of edges in the original network that connect vertices in community i with vertices in community $j = 1, \cdots, K$ and $i \neq j$. Given this, the modularity of $\boldsymbol{C}$ is defined by

$$mod(\boldsymbol{C}) = \sum_{k=1}^{K} [f_{kk}(\boldsymbol{C}) - f_{kk}^{*}]^2 \qquad (1)$$

where f_{kk} is the fraction of edges which connect vertices within the same community k in G, and f_{kk}^{*} is the expected value of f_{kk} under a random edge assignment. Modularity is obtained by maximizing Eq. (1) where the observed fraction of edges is substantially different with the fraction of edges formed via random process. That is, large value in modularity indicates a substantial connection among some vertices than expected, and this suggests the presence of a nontrivial community structure in the network. In practice, the community detection in our PWIDs network was conducted with "fastgreedy.community" algorithm in "igraph" package in R.

2.6 Identification of causal effects with IPW estimator

With notations introduced above, we describe four different causal estimands of interest in the presence of interference (5, 7). As mentioned above, to define the potential outcomes, we assume Bernoulli individual group assignment strategy under the strategy individuals within community i assigned treatment at random with probability α (6). Then, the probability of community i's exposure vector is

$$\pi_i(\boldsymbol{A_i}; \alpha) = \prod_{j=1}^{n_i} \alpha^{A_{ij}} (1-\alpha)^{1-A_{ij}} \quad (2)$$

and the probability of community i's exposure vector which exclude jth individual is

$$\pi_i(\boldsymbol{A_{i,-j}}; \alpha) = \prod_{k=1,k\neq j}^{n_i} \alpha^{A_{ik}} (1-\alpha)^{1-A_{ik}}. \quad (3)$$

When an individual exposure $a \in \{a=0, a=1\}$ with probability α, its average potential outcome is denoted by

$$\bar{Y}_{ij}(a;\alpha) = \sum_{\mathrm{a}_{i,-j} \in \mathcal{A}(n_i-1)} Y_{ij}(a, \mathrm{a}_{i,-j})\, \pi_i(\mathrm{a}_{i,-j}; \alpha). \quad (4)$$

Also, the marginal individual average potential outcome is defined by

$$\bar{Y}_{ij}(\alpha) = \sum_{\mathrm{a}_i \in \mathcal{A}(n_i)} Y_{ij}(\mathrm{a}_i)\, \pi_i(\mathrm{a}_i; \alpha). \quad (5)$$

With this notation, community-level average potential outcome is

$$\bar{Y}_i(a;\alpha) = \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{Y}_{ij}(a;\alpha). \quad (6)$$

Then, population-level average potential outcome with a certain coverage α is

$$\bar{Y}(a;\alpha) = \frac{1}{K} \sum_{i=1}^{K} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{Y}_{ij}(a;\alpha) \right\} = \sum_{i=1}^{K} \bar{Y}_i(a;\alpha). \quad (7)$$

As in the case of the marginal individual average potential outcome, we can express the population average potential outcome with

$$\bar{Y}(\alpha) = \frac{1}{K} \sum_{i=1}^{K} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{Y}_{ij}(\alpha) \right\}. \quad (8)$$

Given these, the following notations represent four different causal effects as proposed in Hudgens and Halloran (5). The *direct* (or individual) effect is defined as:

$$\overline{DE}(\alpha) = \bar{Y}(a = 0; \alpha) - \bar{Y}(a = 1; \alpha), \quad (9)$$

the *indirect* effect is defined as:

$$\overline{IE}(\alpha, \alpha') = \bar{Y}(a = 0; \alpha) - \bar{Y}(a = 0; \alpha'), \quad (10)$$

the *total* effect is defined as:

$$\overline{TE}(\alpha, \alpha') = \bar{Y}(a = 0; \alpha) - \bar{Y}(a = 1; \alpha'), \quad (11)$$

and the *overall* effect is defined as:

$$\overline{OE}(\alpha, \alpha') = \bar{Y}(\alpha) - \bar{Y}(\alpha'), \quad (12)$$

where these causal effects are described as population averages. Tchetgen Tchetgen and VanderWeele proposed an inverse probability weighting (IPW) method to estimate these four different causal effects under the presence of interference in observational studies (6). The IPW estimators are unbiased when the group level propensity scores are known, and the following assumptions hold:

(1) Conditional independence: $\Pr(\boldsymbol{A_i} = \mathrm{a}_i | \boldsymbol{X_i}, \boldsymbol{Y_i}(\cdot)) = \Pr(\boldsymbol{A_i} = \mathrm{a}_i | \boldsymbol{X_i})$;

(2) Positivity: $\Pr(\boldsymbol{A_i} = \mathrm{a}_i | \boldsymbol{X_i}) > 0 \; \forall \mathrm{a}_i \in \mathcal{A}(n_i)$.

In practice, however, the true propensity scores are often unknown, and we need to estimate with

$$f_{\boldsymbol{A_i}|\boldsymbol{X_i}}(\boldsymbol{A_i}|\boldsymbol{X_i}; \theta_x, \theta_s) = \int \prod_{j=1}^{n_i} h_{ij}(b_i; \theta_x)^{A_{ij}} \{1 - h_{ij}(b_i; \theta_x)\}^{1-A_{ij}} f_b(b_i; \theta_s) db_i \quad (13)$$

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 jth individual in community i and $f_b(\cdot; \theta_s)$: the density of community specific random effect b_i which follows a normal distribution with mean 0 and variance θ_s.

With the estimated cluster-level propensity score, the IPW estimator for community-level average potential outcomes is calculated by

$$\hat{Y}_i^{ipw}(a,\alpha) = \frac{\sum_{j=1}^{n_i} \pi_i(\boldsymbol{A}_{i,-j};\alpha) I(A_{ij}=a) Y_{ij}}{n_i f_{\boldsymbol{A}_i|\boldsymbol{X}_i}(\boldsymbol{A}_i|\boldsymbol{X}_i;\widehat{\theta})}, \quad (14)$$

and the marginal potential outcomes is

$$\hat{Y}_i^{ipw}(\alpha) = \frac{\sum_{j=1}^{n_i} \pi_i(\boldsymbol{A}_i;\alpha) I(\boldsymbol{A}_i;a) Y_{ij}}{n_i f_{\boldsymbol{A}_i|\boldsymbol{X}_i}(\boldsymbol{A}_i|\boldsymbol{X}_i;\widehat{\theta})}. \quad (15)$$

Then, population-level IPW estimators of four different causal effects are represented by

$$\widehat{DE}(\alpha) = \hat{Y}^{ipw}(a=0;\alpha) - \hat{Y}^{ipw}(a=1;\alpha) \quad (16)$$

$$\widehat{IE}(\alpha,\alpha') = \hat{Y}^{ipw}(a=0;\alpha) - \hat{Y}^{ipw}(a=0;\alpha') \quad (17)$$

$$\widehat{TE}(\alpha,\alpha') = \hat{Y}^{ipw}(a=0;\alpha) - \hat{Y}^{ipw}(a=1;\alpha') \quad (18)$$

$$\widehat{OE}(\alpha,\alpha') = \hat{Y}^{ipw}(\alpha) - \hat{Y}^{ipw}(\alpha') \quad (19)$$

We used SAS 9.4 (Cary, NC; https://www.sas.com/) for data preparation and R version 3.4.4 (https://cran.r-project.org/) for the visualization and analysis. For the causal inference under the presence of interference, we use newly developed "inferference" package in R that enable us to conduct previously mentioned IPW method (11).

List of References

1. Friedman SR, Curtis R, Neaigus A, Jose B, Des Jarlais DC. Social networks, drug injectors' lives, and HIV/AIDS: *Springer Science & Business Media*; 2006.

2. Andrich D. Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement*. 1978 Oct;2(4):581-94.

3. Likert R, Roslow S, Murphy G. A simple and reliable method of scoring the Thurstone attitude scales. *Personnel Psychology*. 1993 Sep 1;46(3):689-90.

4. Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology*. 1991:331-8.

5. Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. *Statistical Methods in Medical Research*. 2012;21(1):55-75.

6. Hudgens MG, Halloran ME. Toward Causal Inference with Interference. *J Am Stat Assoc*. 2008;103(482):832-42.

7. Sobel ME. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association*. 2006;101(476):1398-407.

8. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*. 2001 Aug;27(1):415-44.

9. Kolaczyk ED, Csárdi G. Statistical analysis of network data with R: *Springer*; 2014.

10. Newman M. Networks: an introduction: *Oxford University Press*; 2010.

11. Saul BC, Hudgens MG. A Recipe for inferference: Start with Causal Inference. Add Interference. Mix Well with R. *Journal of Statistical Software*. 2017;82(1):1-21.

CHAPTER 3

RESULTS

3.1 Descriptive statistics

Table 2 is a descriptive summary of participants' characteristics in the SFHR PWIDs network. Among 402 participants, 40.9% were HIV infected, 5.1% developed AIDS and 75.1% had hepatitis B virus (HBV) infected. Other basic demographics such as gender, race/ethnicity, highest education, current job status and the place currently live are shown in Table 2. About 64.3% of participants had less than 12^{th} grade and 90.8% of participants were unemployed. To investigate internal consistency of responses in SFHR health belief questions (Table 1), we calculated Cronbach's alpha for health belief and blame relates questions, respectively, on their original integer scale. In calculating the reliability measure with 402 participants, the values of alpha for health belief and blame questions were 0.59 and 0.14, respectively. These values are far from the acceptance for internal consistency. However, in calculating the Cronbach's alphas with 767 participants, the values of alpha for health belief and blame questions were 0.77 and 0.47, respectively. This indicates the acceptable internal consistency in health belief questions, though the internal consistency in blame was still questionable. Further exploration regarding the reliability and validity is made in discussion section.

Table 3 and 4 show the relationship between exposures (i.e. locus of control, blame attitudes) and outcomes (i.e. receipt of HIV testing result, recent medical visit)

without controlling covariates. Table 3 gives the contingency tables of showing the relationship between locus of control and health-seeking behaviors and those tables were created on both people who belong to the giant component and people do not in the giant component. Table 3 also include odds ratios calculated based on the information obtained from the contingency tables. As for the relation between locus of control and receipt of HIV testing result, all odds ratios were greater than one and indicated that those with internal locus of control are more likely to receive a HIV testing result. However, all the corresponding 95% CIs included the value of one, so the relationship between locus of control and receipt of a HIV testing result was not statistically significant. The same result was observed when we saw the relationship between blame attributes and receipt of a HIV testing result (Table 4). As for the relation between locus of control and recent medical visit, odds ratios were greater that one except among those who are not in the giant component. However, all 95% CIs include the value of one, we could not observe any significant relationship between locus of control and recent medical visit. The same result was observed when we saw the relationship between blame attributes and recent medical visit (Table 4). In short, without neither considering dissemination effect or controlling covariates, we could not observe any significant relationship between attitude toward HIV/AIDS risk and health-seeking behaviors among PWIDS in SFHR study.

Table2. Descriptive Summary Table of Participants' Characteristics in the Social Factors and HIV Risk Study from Bushwick, New York (1991-1993) (n = 402).

Characteristics		Number of participants (%)
Sex	Male	287 (71.4%)
	Female	115 (28.6%)
Age	Young adult (18-40 years)	290 (72.1%)
	Middle-aged (>40 years)	112 (27.9%)
Race/ethnicity	White	153 (38.1%)
	Others	249 (61.9%)
Highest education	Less than High School Graduation (or LT 12^{th} grade)	258 (64.3%)
	High School or more	143 (35.7%)
Work status	No Job	364 (90.8%)
	Some work	37 (9.2%)
Where currently live	In your own apartment or house	116 (28.9%)
	Someone else's apartment or house	192 (47.8%)
	Homeless/others	94 (23.4%)
Number of participants with HIV data (% positive)		396 (40.9%)
Number of participants with HBV data (% positive)		325 (75.1%)
Number of participants with AIDS data (% positive)		395 (5.1%)

HIV: human immunodeficiency virus
HBV: hepatitis B virus
AIDS: acquired immunodeficiency syndrome

Table 3. Observed Relationships between Locus of Control and Health-Seeking Behaviors without Controlling Covariates and Corresponding Odds Ratios with their 95% CI in the Social Factors and HIV Risk Study from Bushwick, New York (1991-1993) (n = 402).

	Giant		Not Giant		Odds Ratio (95% CI)		
	Internal	External	Internal	External	Giant	Not Giant	Total
Received	23	3	46	5	1.94 (0.55, 6.85)	1.64 (0.59, 4.57)	1.87 (0.85, 4.11)
Not received	138	35	129	23			
Recent medical visit	141	32	147	24	1.32 (0.49, 3.56)	0.88 (0.28, 2.72)	1.07 (0.51, 2.24)
No recent medical visit	20	6	28	4			

Giant: giant component
CI: confidence interval

Table 4. Observed Relationships between Blame attributes and Health-Seeking Behaviors without Controlling Covariates and Corresponding Odds Ratios with their 95% CI in the Social Factors and HIV Risk Study from Bushwick, New York (1991-1993) (n = 402).

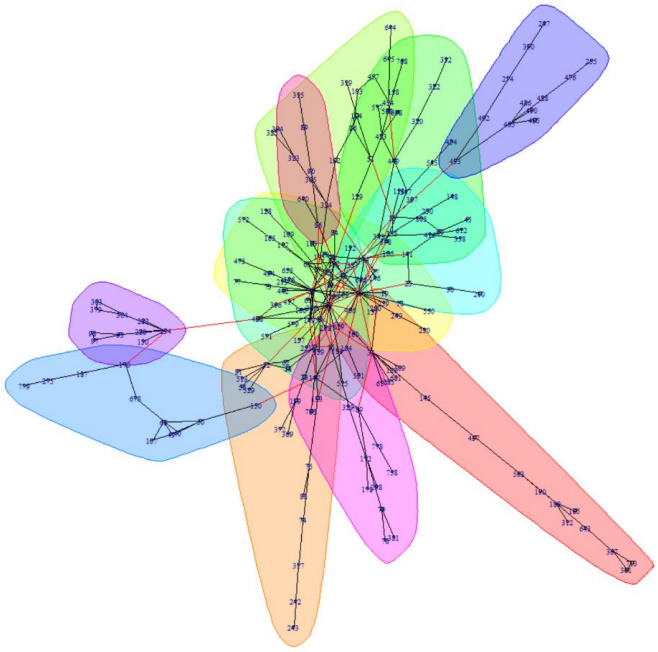
	Giant		Not Giant		Odds Ratio (95% CI)		
	Self-blame	Others	Self-blame	Others	Giant	Not Giant	Total
Received	18	8	34	17	1.35 (0.56, 3.29)	1.00 (0.52, 1.98)	1.15 (0.68, 1.96)
Not received	108	65	101	51			
Recent medical visit	110	63	113	58	1.09 (0.47, 2.55)	0.89 (0.39, 1.99)	0.96 (0.54, 1.74)
No recent medical visit	16	10	22	10			

Giant: giant component
CI: confidence interval

3.2 Community detection

In the entire SFHR network (Figure 2), there is 85 connected components and one of them forms the giant component that include 199 participants. To find reasonable communities of PWIDs, first we considered each of 84 connected components except the giant component forms its own community, then, we conducted community detection within the giant component to find smaller communities in which members were more closely connected. This approach allows for differences between the information about connections within the communities that belong to the giant component and the rest of the communities. For example, we do not know whether two different communities are connected either closely, distantly or not connected at all because we do not have any information of shared risk links among connected components outside of the giant component. As the result of conducting community detection, we obtained 12 communities in the giant component (Figure 5). Based on this process, we obtained 96 communities of PWIDs in total. Of the 96 communities, the smallest communities consisted of 2 participants (there were 66 smallest communities), and the largest community included 35 participants. The average order or the number of participants included in a community was 4.2 participants.

Figure 5. Result from Community Detection against the giant component in the Social Factors and HIV Risk Study PWIDs' Network with 12 Communities in the Giant Component.



3.3 Causal effects estimation

Causal effects under the presence of dissemination were defined by comparing community-level potential outcomes with different coverages of exposure, which is the proportion of people who have internal locus of control/self- blame attribute in a community. Given the 96 communities of PWIDs identified by community detection, the observed distributions of coverage are shown in Figure 5. The left panel shows the frequency distribution of observed coverage or proportion of people who have internal locus and the right panel shows the frequency distribution of coverage or proportion of people who have self-blame attributes in each community out of ninety-six. The coverage of self-blame attribute has wider variation than that of internal locus. Also, because the entire SFHR PWIDs networks included 66 communities of PWIDS that have only two participants as its member, we observed many communities with the

coverage of 0%, 50% and 100%. As sufficient number of communities were observed to estimate group-level propensity score, we focused on 50%, 70% and 99% coverages for causal estimation. There were four different models because we considered two different exposures (i.e. locus of control and blame attitude) and two different health seeking behaviors (i.e. receipt of HIV testing result and medical visit within the past year). Sex, race (White, Others), education level (less than high school, high school or more), age (young adult (18-40 years), middle-aged (>40 years)) and all pairwise interaction terms were used as individual-level covariates in the models.

Figure 6. Observed Proportion of Participants with Internal Locus of Control per Community (Left) and Observed Proportion of Individuals with Self-Blame Attribute per Community (Right).

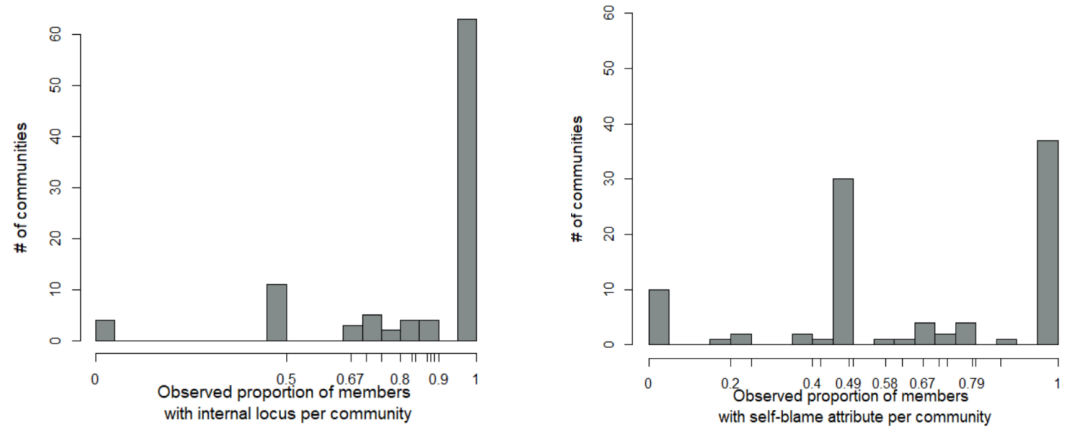


Table 5 represents the estimation results of four different causal estimands and corresponding 95% confidence intervals (CIs) of the effect of locus of control on receipt of HIV testing result, including the results from unadjusted, adjusted and adjusted with interactions models. Figure 7 shows the plots of estimates four different causal effects in this model. Coverage is defined as the proportion of people with

internal locus of control in a community, $\alpha < \alpha'$. For both 50% and 70% coverage groups, the direct effects (adjusted with interactions) were statistically significant and those with external locus of control was a 16% less likely to receive their HIV testing result compare to those with internal locus of control in those coverage groups (95% confidential intervals (CIs): -0.265, -0.055 for 50% coverage; -0.268, -0.055 for 70% coverage, respectively). The indirect effect was significant only when comparing 50% and 70% coverage groups. This means that when the coverage of people with internal locus was decreased from 70% to 50%, the likelihood of receiving the HIV testing result decreased about 3% for those with external locus (95% CI: -0.054, - 0.008). The total effects, that is, the maximal impact of locus of control, were statistically significant in the comparison of all coverage groups. There were about 19% reduction in the likelihood of receiving their HIV test result between for those with external locus with 50% coverage compared to those with internal in 70% or 99% coverage groups (95% CIs: -0.286, -0.100 for 50% vs. 70%; -0.291, -0.093 for 50% vs. 99%, respectively). A 16% difference in total effect was observed in the comparison between those with external in 70% and those with internal in 99% groups (95%CI: -0.272, -0.049). The estimated overall effects support that the marginal likelihood of receipt of HIV testing result is significantly lower in groups with low coverage of internal compare to high coverage group. For example, the likelihood of receipt is about 7% lower in 50% coverage group compare to 70% coverage group, while the likelihood is about 11% lower in 50% coverage group compare to 99% coverage group (95% CI: -0.089, -0.041 for 50% vs. 70%; -0.190, -0.032 for 50% vs. 99%, respectively).

Table 5. Unadjusted and Adjusted Estimated Risk Differences (RDs) with Corresponding 95% Confidence Intervals (95% CIs) of Causal Effects of Locus of Control (External vs. Internal) on Likelihood of Receiving SFHR HIV Test Results among 402 Participants in the Social Factors and HIV Risk Study from Bushwick, New York (1991 - 1993).

		Unadjusted		Adjusted		Adjusted with interactions	
Effect	Coverage (α , α')	RD	95%CI	RD	95%CI	RD	95%CI
Direct	(50%, 50%)	**-0.148**	(-0.230, -0.065)	**-0.159**	(-0.251, -0.066)	**-0.160**	(-0.265, -0.055)
Direct	(70%, 70%)	**-0.142**	(-0.246, -0.038)	**-0.163**	(-0.261, -0.065)	**-0.162**	(-0.268, -0.055)
Direct	(99%, 99%)	-0.101	(-0.258, 0.056)	-0.118	(-0.265, 0.029)	-0.130	(-0.268, 0.008)
Indirect	(50%, 70%)	**-0.041**	(-0.071, -0.012)	**-0.032**	(-0.058, -0.007)	**-0.031**	(-0.054, -0.008)
Indirect	(50%, 99%)	-0.070	(-0.1560, 0.019)	-0.072	(-0.152, 0.009)	-0.062	(-0.123, 0.000)
Indirect	(70%, 99%)	-0.029	(-0.098, 0.040)	-0.039	(-0.097, 0.018)	-0.030	(-0.072, 0.011)
Total	(50%, 70%)	**-0.183**	(-0.271, -0.096)	**-0.195**	(-0.282, -0.109)	**-0.193**	(-0.286, -0.100)
Total	(50%, 99%)	**-0.172**	(-0.278, -0.066)	**-0.190**	(-0.289, -0.092)	**-0.192**	(-0.291, -0.093)
Total	(70%, 99%)	**-0.130**	(-0.254, -0.006)	**-0.158**	(-0.269, -0.046)	**-0.161**	(-0.272, -0.049)
Overall	(50%, 70%)	**-0.067**	(-0.096, -0.038)	**-0.067**	(-0.093, -0.041)	**-0.065**	(-0.089, -0.041)
Overall	(50%, 99%)	**-0.097**	(-0.183, -0.010)	**-0.109**	(-0.190, -0.029)	**-0.111**	(-0.190, -0.032)
Overall	(70%, 99%)	-0.030	(-0.095, 0.035)	-0.042	(-0.104, 0.020)	-0.046	(-0.105, 0.013)

* The adjusted model with interaction terms includes all pairwise interactions of sex and race, education, and age.

** Coverage is defined as the proportion of people with internal locus of control in a community. $\alpha < \alpha'$.

Figure 7. Plots of IPW Estimates for Four Different Causal Effects of Locus of Control on Receipt of HIV Testing Result. Direct Effect (top-left), Indirect Effect (top-right), Total Effect (bottom-left) and Overall Effect (bottom-right). The polygon represents 95% CI.

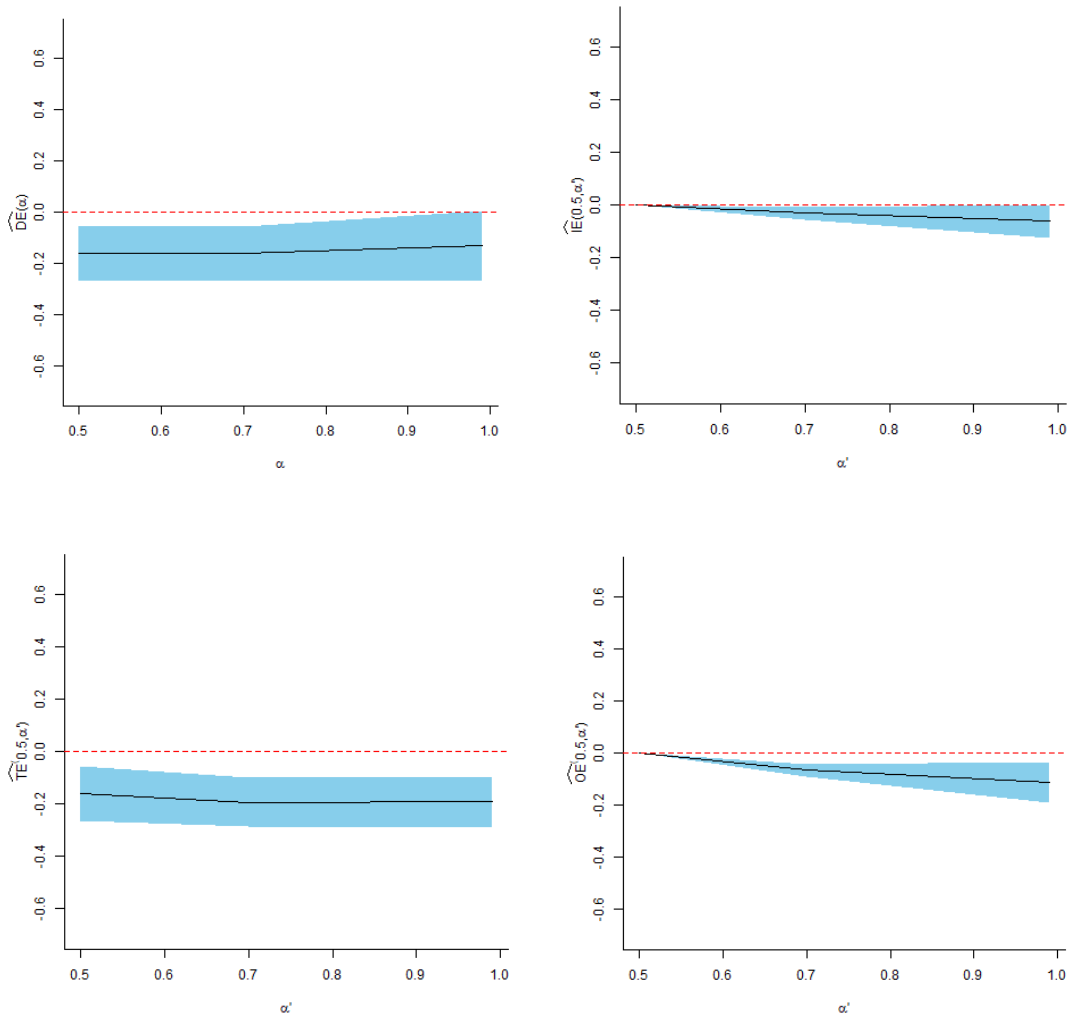


Table 6 represents the estimation results of four different causal estimands and corresponding 95% confidence intervals (CIs) of the effect of locus of control on recent medical encounters, including the results from unadjusted, adjusted and adjusted with interactions models. Figure 8 shows the plots of estimates four different causal effects in this model. Coverage is defined as the proportion of people with internal locus of control in a community, $\alpha < \alpha'$. A 28 % significant direct effect was observed in 99% coverage group (95%CI: -0.470, -0.089). This means that in 99% coverage of internal locus, people with external locus are 28% less likely to go picking up HIV testing result compare to those with internal locus. Unlike the results obtained from the effect on receipt of HIV testing result, there were no statistically significant effects in indirect and total effects in all coverage groups. However, the estimated overall effects show that the marginal likelihood of having recent medical encounters is significantly lower in low coverage group compare to high coverage one. The likelihood of visit a doctor is about 25% lower in 50% coverage group compare to 99% coverage group, while the likelihood is about 11% lower in 70% coverage group compare to 99% coverage group (95%CIs: -0.430, -0.061 for 50% vs. 99%; -0.181, -0.047 for 70% vs. 99%, respectively).

Table 6. Unadjusted and Adjusted Estimated Risk Differences with Corresponding 95% Confidence Intervals (95% CIs) of Causal Effects of Locus of Control (External vs. Internal) on Likelihood of Reporting a Doctor's Visit in the Last Year among 402 Participants in the Social Factors and HIV Risk Study from Bushwick, New York (1991 - 1993).

Effect	Coverage (α , α')	Unadjusted		Adjusted		Adjusted with interactions	
		RD	95%CI	RD	95%CI	RD	95%CI
Direct	(50%, 50%)	0.211	(-0.286, 0.708)	0.229	(-0.255, 0.713)	0.090	(-0.271, 0.451)
Direct	(70%, 70%)	0.003	(-0.296, 0.301)	-0.024	(-0.334, 0.286)	-0.111	(-0.346, 0.123)
Direct	(99%, 99%)	-0.227	(-0.463, 0.009)	**-0.252**	(-0.468, -0.035)	**-0.280**	(-0.470, -0.089)
Indirect	(50%, 70%)	-0.001	(-0.260, 0.257)	0.048	(-0.177, 0.272)	-0.008	(-0.181, 0.165)
Indirect	(50%, 99%)	0.208	(-0.298, 0.715)	0.206	(-0.293, 0.705)	0.077	(-0.311, 0.464)
Indirect	(70%, 99%)	0.210	(-0.063, 0.482)	0.158	(-0.127, 0.443)	0.085	(-0.136, 0.305)
Total	(50%, 70%)	0.001	(-0.537, 0.539)	0.024	(-0.489, 0.536)	-0.119	(-0.496, 0.258)
Total	(50%, 99%)	-0.019	(-0.491, 0.453)	-0.046	(-0.527, 0.435)	-0.203	(-0.559, 0.153)
Total	(70%, 99%)	-0.017	(-0.255, 0.220)	-0.093	(-0.374, 0.187)	-0.195	(-0.409, 0.018)
Overall	(50%, 70%)	-0.105	(-0.316, 0.106)	-0.084	(-0.268, 0.101)	-0.131	(-0.265, 0.003)
Overall	(50%, 99%)	-0.122	(-0.358, 0.114)	-0.158	(-0.405, 0.089)	**-0.246**	(-0.430, -0.061)
Overall	(70%, 99%)	-0.017	(-0.082, 0.048)	-0.074	(-0.154, 0.006)	**-0.114**	(-0.181, -0.047)

* The adjusted model with interaction terms includes all interactions of sex and race, education, and age.

** Coverage is defined as the proportion of people with internal locus of control in a community. $\alpha < \alpha'$.

Figure 8. Plots of IPW Estimates for Four Different Causal Effects of Locus of Control on a Medical Encounter within the Past Year. Direct Effect (top-left), Indirect Effect (top-right), Total Effect (bottom-left) and Overall Effect (bottom-right). The polygon represents 95% CI.

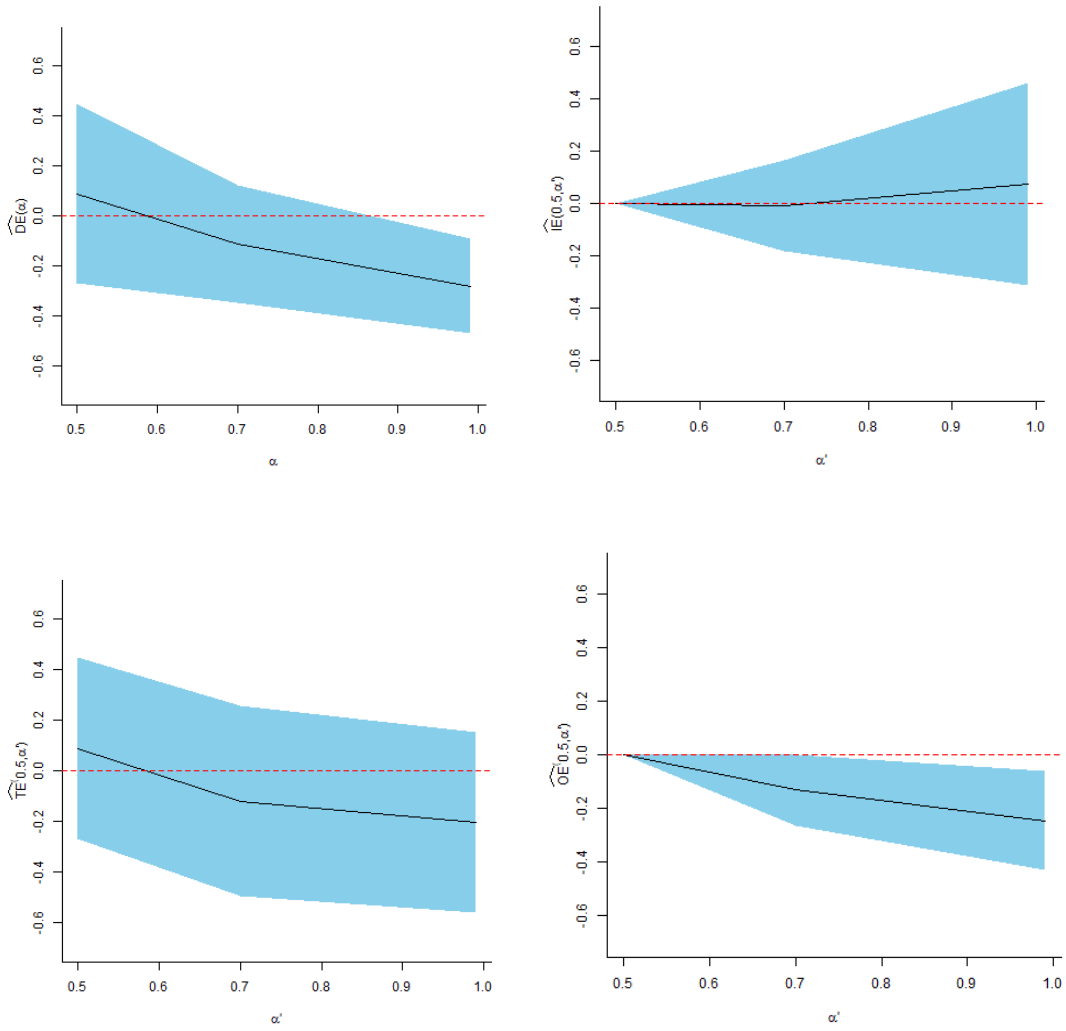


Table 7 represents the estimation results of four different causal estimands and corresponding 95% confidence intervals (CIs) of the effect of blame attribute on receipt of HIV testing result, including the results from unadjusted, adjusted and adjusted with interactions models. Figure 8 shows the plots of estimates four different causal effects in this model. Coverage is defined as the proportion of people with self-blame attribute in a community, $\alpha < \alpha'$. As the reminder, the coverage here is defined as the proportion of people who have self-blame attribute in a community. As you can see from Table 7, none of the estimates were statistically significant. Figure 9 also shows 95% CIs for all estimate include zero in this model. Therefore, we could not determine the direction of impact of PWIDs' individual blame attribute on their receipt of HIV testing results from al the estimation of four different causal effects.

Table 7. Unadjusted and Adjusted Estimated Risk Differences with Corresponding 95% Confidence Intervals (95% CIs) of Causal Effects of Blame (Others vs. Self) on Likelihood of Receiving SFHR HIV Test Results among 402 Participants in the Social Factors and HIV Risk Study from Bushwick, New York (1991 - 1993).

Effect	Coverage (α, α')	Unadjusted		Adjusted		Adjusted with interactions	
		RD	95%CI	RD	95%CI	RD	95%CI
Direct	(50%, 50%)	-0.043	(-0.164, 0.079)	-0.059	(-0.176, 0.059)	-0.045	(-0.159, 0.069)
Direct	(70%, 70%)	-0.035	(-0.159, 0.088)	-0.040	(-0.163, 0.083)	-0.034	(-0.154, 0.085)
Direct	(99%, 99%)	-0.077	(-0.250, 0.096)	-0.071	(-0.245, 0.102)	-0.065	(-0.230, 0.100)
Indirect	(50%, 70%)	-0.002	(-0.047, 0.044)	-0.007	(-0.050, 0.037)	-0.001	(-0.043, 0.041)
Indirect	(50%, 99%)	0.034	(-0.077, 0.145)	0.024	(-0.082, 0.131)	0.034	(-0.066, 0.134)
Indirect	(70%, 99%)	0.036	(-0.037, 0.109)	0.031	(-0.040, 0.102)	0.035	(-0.032, 0.102)
Total	(50%, 70%)	-0.037	(-0.164, 0.090)	-0.047	(-0.170, 0.077)	-0.035	(-0.154, 0.083)
Total	(50%, 99%)	-0.043	(-0.200, 0.115)	-0.047	(-0.201, 0.107)	-0.031	(-0.175, 0.113)
Total	(70%, 99%)	-0.041	(-0.197, 0.115)	-0.040	(-0.196, 0.115)	-0.030	(-0.176, 0.116)
Overall	(50%, 70%)	-0.005	(-0.048, 0.038)	-0.005	(-0.047, 0.036)	-0.003	(-0.042, 0.036)
Overall	(50%, 99%)	-0.021	(-0.135, 0.094)	-0.017	(-0.129, 0.095)	-0.008	(-0.113, 0.097)
Overall	(70%, 99%)	-0.015	(-0.095, 0.064)	-0.012	(-0.091, 0.067)	-0.005	(-0.079, 0.068)

* The adjusted model with interaction terms includes all interactions of sex and race, education, and age.

** Coverage is defined as the proportion of people with self-blame attribute in a community. $\alpha < \alpha'$.

Figure 9. Plots of IPW Estimates for Four Different Causal Effects of Blame Attribute on Receipt of HIV Testing Result. Direct Effect (top-left), Indirect Effect (top-right), Total Effect (bottom-left) and Overall Effect (bottom-right). The polygon represents 95% CI.

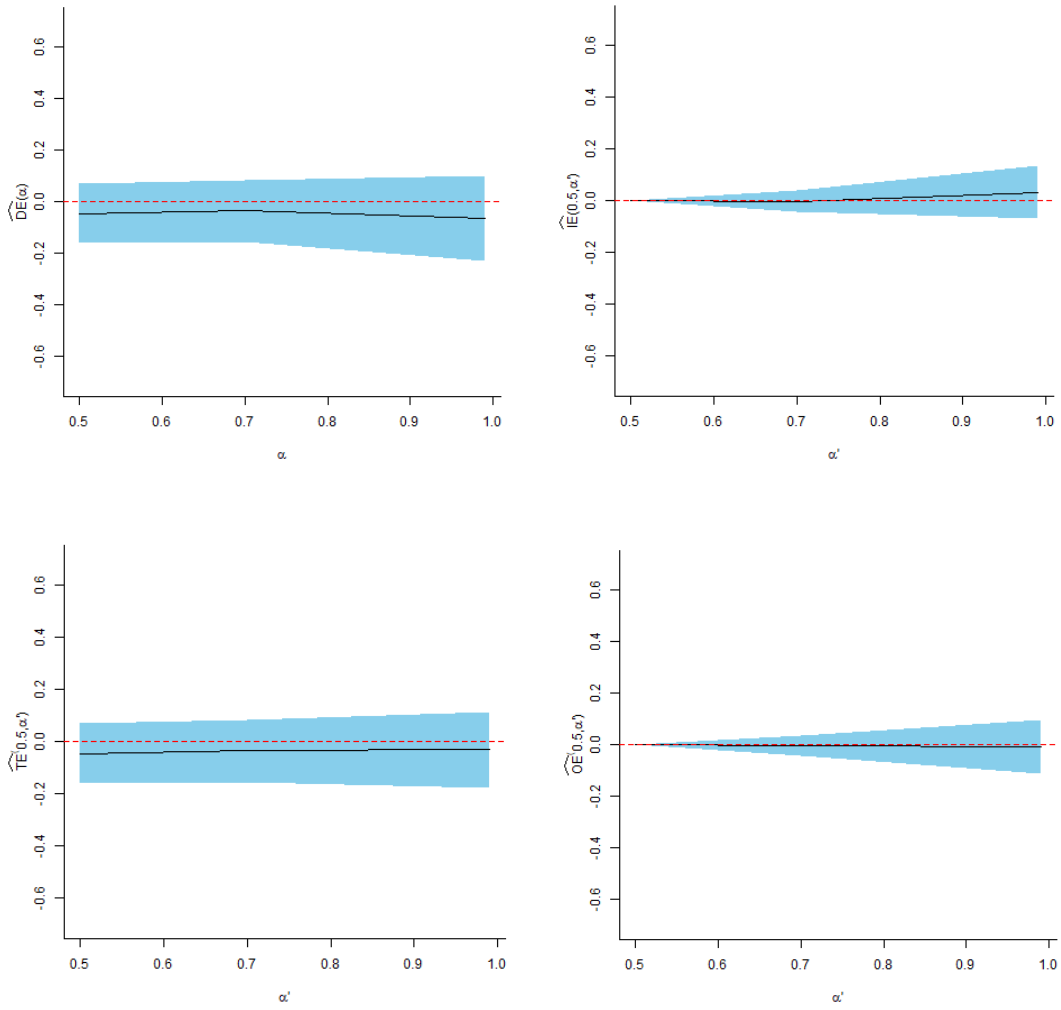


Table 8 represents the estimation results of four different causal estimands and corresponding 95% confidence intervals (CIs) of the effect of the blame attribute on recent medical encounters, including the results from unadjusted, adjusted and adjusted with interactions models. Figure 10 shows the plots of estimates four different causal effects in this model. Coverage is defined as the proportion of people with self-blame attribute in a community, $\alpha < \alpha'$. We observed statistically significant results for both the direct and indirect effects. About 27% significant direct effect was observed in 99% coverage group (95% CI: -0.527, -0.011). This can be interpreted as, in the 99% coverage group, people who blame others are 27% less likely to have a recent medical encounter compare to those who blame themselves. More interesting result was observed in indirect effect. That is, when the coverage for those who blame others decreased, there was an increase of likelihood of recent medical encounters. For example, when the coverage of people with self-blame attribute was decreased from 99% to 50%, the likelihood of recent medical encounters increased about 27% for those with blame-others attribute (95% CI: 0.072, 0.472), and when the coverage of people with self-blame attribute was decreased from 99% to 70%, the likelihood of recent medical encounters increased about 19% for those with blame others attribute (95% CI: 0.038, 0.335).

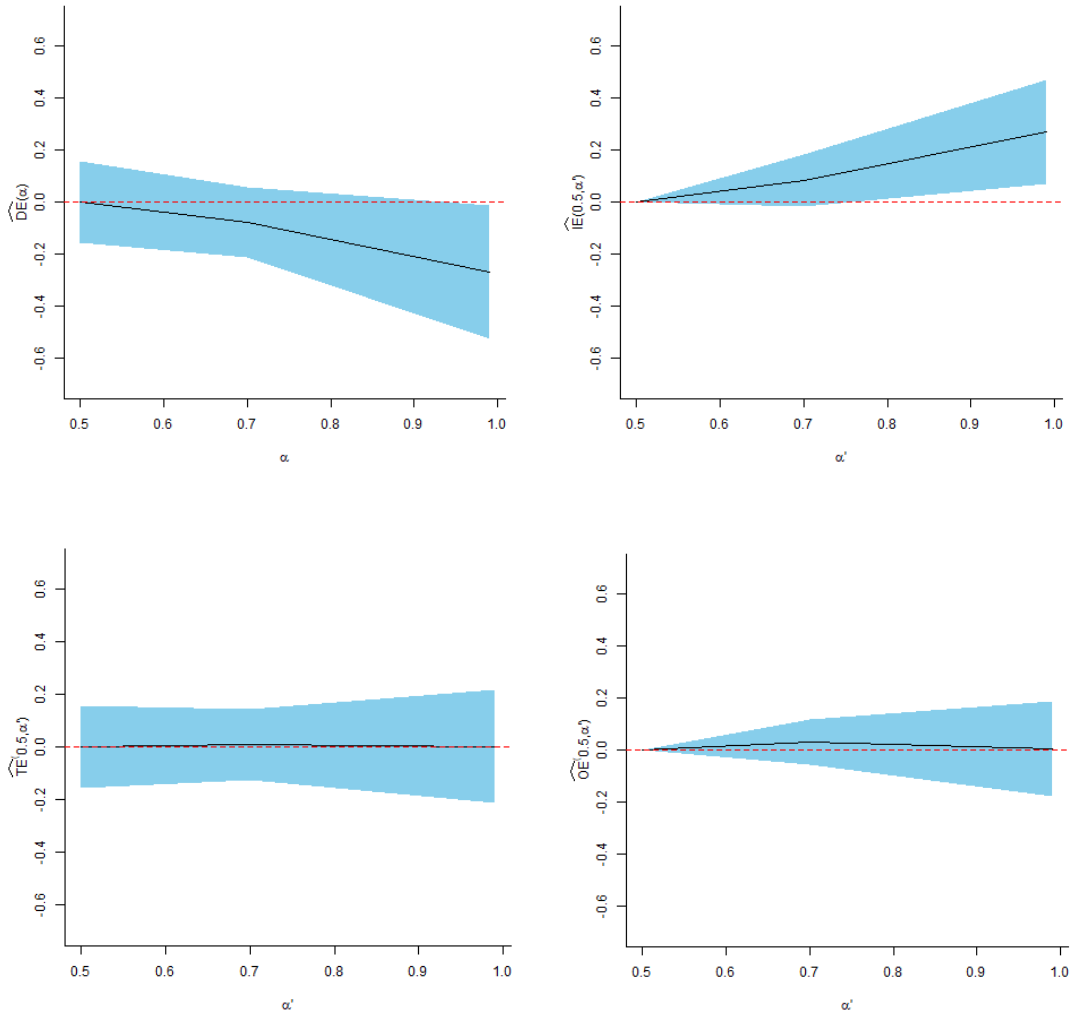
Table 8. Unadjusted and Adjusted Estimated Risk Differences with Corresponding 95% Confidence Intervals (95% CIs) of Causal Effects of Blame (Others vs. Self) on Likelihood of Reporting a Doctor's Visit in the Last Year among 402 Participants in the Social Factors and HIV Risk Study from Bushwick, New York (1991 - 1993).

Effect	Coverage (α, α')	Unadjusted		Adjusted		Adjusted with interactions	
		RD	95%CI	RD	95%CI	RD	95%CI
Direct	(50%, 50%)	0.017	(-0.141, 0.175)	0.014	(-0.158, 0.185)	0.002	(-0.156, 0.160)
Direct	(70%, 70%)	-0.054	(-0.205, 0.096)	-0.062	(-0.218, 0.095)	-0.076	(-0.213, 0.060)
Direct	(99%, 99%)	-0.209	(-0.467, 0.050)	-0.234	(-0.509, 0.041)	**-0.269**	(-0.527, -0.011)
Indirect	(50%, 70%)	0.076	(-0.001, 0.153)	0.069	(-0.008, 0.146)	0.086	(-0.014, 0.186)
Indirect	(50%, 99%)	**0.238**	(0.053, 0.423)	**0.230**	(0.032, 0.428)	**0.272**	(0.072, 0.472)
Indirect	(70%, 99%)	**0.162**	(0.019, 0.305)	**0.161**	(0.011, 0.311)	**0.186**	(0.038, 0.335)
Total	(50%, 70%)	0.022	(-0.099, 0.143)	0.072	(-0.134, 0.148)	0.009	(-0.126, 0.144)
Total	(50%, 99%)	0.029	(-0.159, 0.218)	-0.004	(-0.216, 0.208)	0.003	(-0.214, 0.219)
Total	(70%, 99%)	-0.047	(-0.263, 0.170)	-0.073	(-0.302, 0.155)	-0.083	(-0.299, 0.133)
Overall	(50%, 70%)	0.029	(-0.022, 0.081)	0.019	(-0.030, 0.068)	0.031	(-0.054, 0.117)
Overall	(50%, 99%)	0.023	(-0.119, 0.164)	-0.008	(-0.171, 0.154)	0.005	(-0.178, 0.187)
Overall	(70%, 99%)	-0.007	(-0.148, 0.134)	-0.027	(-0.184, 0.129)	-0.027	(-0.183, 0.130)

* The adjusted model with interaction terms includes all interactions of sex and race, education, and age.

** Coverage is defined as the proportion of people with self-blame attribute in a community. $\alpha < \alpha'$.

Figure 10. Plots of IPW Estimates for Four Different Causal Effects of Blame Attribute on a Medical Encounter within the Past Year. Direct Effect (top-left), Indirect Effect (top-right), Total Effect (bottom-left) and Overall Effect (bottom-right). The polygon represents 95% CI.



CHAPTER 4

DISCUSSION

In this work, we evaluated the effect of PWIDs' individual attitudes toward HIV/AIDS risk on their own and their neighbors' health seeking behaviors, including receipt of HIV testing results and recent medical encounters. First, as for the effect of health beliefs on receipt of HIV testing results, PWIDs with external locus were less likely to receive their HIV testing result in regardless of the proportion of people in their network who reported internal locus of control. More importantly, we observed a possible dissemination of health beliefs. PWIDs with external locus were distinctly less likely to receive their HIV testing result when they belong to a community with less network members with internal. Possibly, PWIDs who do not believe they are in control of whether or not they get HIV (i.e. with external locus of control) may be less motivated to obtain their HIV testing result and not encouraged by others in their community to get tested and receive care. For the direct effect of health beliefs on recent medical encounters, in a community in which most members have internal locus, people with external locus were 28% less likely to have a recent medical encounter compare to those with internal locus. A similar impact was observed for the effect of the blame attribute on recent medical encounters. In 99% coverage group, people who blame others were 27% less likely to receive their HIV testing result compared to those with self-blame attribute. Individuals with external locus/blame others attribute may believe their risk to acquire HIV/AIDS is relatively low when

most of their community members have internal locus/self-blame attributes and take appropriate health-seeking behaviors to prevent HIV/AIDS, rendering action on their part perceived to be less critical. In addition, the most compelling result was substantial positive dissemination effects when the coverage of people with self-blame attribute was decreased. This means that PWIDs who blame others were more likely to have at least one recent medical encounter when they are closely connected to other individuals who blamed others. Though it requires much additional investigation for a complete understanding, these positive dissemination effects could be explained by that PWIDs with blame others attribute may believe that they can have comparatively higher risk to contract HIV/AIDS because they perceive their community members as not engaging in behaviors that support HIV prevention.

We are the first study to evaluate the effects of attitudes of PWIDs on their health-seeking behaviors with taking into account the possibility of dissemination. Results obtained from this study support the existence of dissemination effects of locus of control/blame attitudes in the SFHR PWIDs network. This emphasizes the importance of network-based interventions as a possibly more effective approach to prevent HIV/AIDS among PWIDs. For example, network-targeted educational programs on HIV/AIDS taking dissemination into account could be helpful for PWIDs to disseminate knowledge about HIV/AIDS risk reduction. This may eventually change their attitudes toward HIV/AIDS risk that support PWIDs to take engage in health-seeking behaviors by leveraging dissemination effect.

There are several limitations to this study. First, we should reemphasize that the information about medical visit is self-reported and there remains the possibility of

response bias in this study. Secondly, the reliability and validity of health belief/blame score thoroughly as a measure of PWID's individual attitudes toward HIV/AIDS risk needs to be considered. As mentioned in result section, regarding the reliability, which examines the internal consistency in responses to SFHR questions shown in Table 1, we calculated Cronbach's alpha, and in calculating it with the entire data including 767 participants, the values of alpha for health belief and blame questions were 0.77 and 0.47, respectively. This indicates the acceptable internal consistency in health belief questions, though the internal consistency in blame was questionable. This low value for alpha in blame questions might be attributed to a small number of items to measure blame attribute. Regarding the validity, it is required to consider if our exposures are valid to measure the attitudes (i.e. locus of control and blame attributes) of PWIDs toward HIV/AIDS risk. In our analysis, we considered a three-level scale (i.e. -1, 0, and 1) while the original response has six-level (i.e. strongly agree, agree, somewhat agree, don't know, disagree, and strongly disagree) and then creating binary exposure variables with a certain threshold. Jacoby and Matell (1971) empirically demonstrated that reliability and validity would be unaffected by collapsing categories of Likert scale into a trichotomous measure (1). Therefore, our approach to use three-level scale would be reasonable; however, future studies are needed to evaluate the validity of our scoring system. We selected threshold based on likelihood to create binary exposure outcomes. Because the existing causal inference methods under dissemination can applicable only for a binary exposure, our approach would be the best possible way to evaluate the effects of PWIDs' attitudes on their health-seeking behavior given dissemination. In a future work, to give more rational to the reliability

and validity of individual attitudes in the SFHR network, we will apply item response models as discussed in Andrich (1978) and Samejima (1972) (2, 3) and develop methodology for dissemination of a categorical exposure.

Thirdly, some assumptions in this study could be relaxed in future work. Although we assumed no homophily, it is impossible to measure homophily in this setting because SHRF PWIDs network was observed at a single time point. One feasible sensitivity analysis to check the existence of homophily would be to compare obtained estimation results with simulation results obtained by creating random networks with same number of node, links and node attributes as the original SFHR PWIDs network. This comparison will enable us to assure the assumption of no homophily. That is, when there is no significant difference between the causal estimation results obtained from our study and random network, we can guarantee that there are no other unobserved factors that explain the relationship between an exposure and outcome behaviors except covariates and exposures used in our study. We may apply more realistic treatment allocation strategies than the assumption of Bernoulli individual group assignment strategy. For example, Barkley *et al*. (2017) and Papadogeorgou *et al*. (2017) introduced a new treatment allocation strategy that allows the correlation of treatment assignment of individuals in a same cluster, which is a generalized version of the Bernoulli individual group assignment strategy (4, 5). Considering this correlation seems more realistic in dealing with observational studies because, in our setting, individual attitudes may not be independent if one has a close connection with another person.

Finally, in this study, there is a possibility that the health-seeking behaviors of PWIDs has changed because the SFHR study was conducted in the early 1990s. However, our results may contribute to understanding how PWIDs attitudes and behaviors have changed over a few decades by conducting a similar analysis in more contemporary studies and additionally provide insights into health-seeking behaviors during an emerging infectious disease epidemic.

By applying causal inference method in the presence of dissemination, we evaluated how personal health attitudes impact health-seeking behavior among PWIDs and their HIV risk networks with the Social Factors and HIV Risk Study (SFHR). Our findings that PWIDs' attitudes affect health seeking behaviors of other members in the same community indicates that interventions that taking a network structure among PWIDs and the existence of dissemination into account can be a more effective and powerful approach to prevent HIV/ADIS transmission among injection drug users. As more concrete suggestion, *segmentation* approach, a type of network interventions in which certain groups of people will be targeted for intervention to bring about behavioral change, may be effective in our setting (6). For example, if we could conduct an educational program, which enable people to come to have internal locus, against the members belong to 50% coverage group, their likelihood to receive HIV testing results would increase about 20% in terms of total effect. Thus, our results could not only support the effectiveness of network interventions and provide an important information to improve HIV prevention interventions among PWIDs.

List of References

1. Jacoby J, Matell MS. Three-point Likert scales are good enough. *Journal of marketing research*. 1971 Nov 1;8(4):495-500.

2. Andrich D. Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement*. 1978 Oct;2(4):581-94.

3. Samejima F. A general model for free-response data. *Psychometrika Monograph Supplement.* 1972 Mar.

4. Barkley BG, Hudgens MG, Clemens JD, Ali M, Emch ME. Causal Inference from Observational Studies with Clustered Interference. *arXiv preprint arXiv:171104834*. 2017.

5. Papadogeorgou G, Mealli F, Zigler C. Causal inference for interfering units with cluster and population level treatment allocation programs. arXiv preprint arXiv:1711.01280. 2017 Nov 3.

6. Valente TW. Network interventions. *Science*. 2012 Jul 6;337(6090):49-53.

APPENDICES

Table A-1. Contingency Tables for All the Pairs of Characteristic Variables of Individual Participants in the Social Factors and HIV Risk Study from Bushwick, New York (1991 - 1993).

		Sex		Race		Education		Work		Live		
		Female	Male	White	Others	LT HS	HS more	No	Some	Own	Someone	Homeless
Sex	Female											
	Male											
Race	White	46	107									
	Others	69	180									
Education	LT HS	74	185	90	169							
	HS more	41	101	63	80							
Work	No	105	260	137	228	238	127					
	Some	10	27	16	21	21	16					
Live	Own	43	73	41	75	60	56	102	14			
	Someone	45	147	72	120	133	59	175	17			
	Homeless	27	67	40	54	66	28	88	6			
Age	Young	89	201	109	68	195	95	267	23	72	150	68
	Middle	26	86	44	181	64	48	98	14	44	42	26

* The null hypothesis of Chi-square test of independence was rejected between Sex and Live, Education and Live, and Age and Live (p-value: 0.037, 0.002, 0.009, respectively).

Table A-1 shows contingency tables of a pair of participants' characteristic variables. Chi-squared test of independence showed that there is statistically significant dependence between sex and living place and education and living place.

Table A-2. Comparison of descriptive network characteristics between SFHR PWID network used in analysis in this study (402 enrolled participants and 403 shared risk links) and that include all risk connections collected in SFHR study (767 enrolled participants, 2,498 disenrolled subjects and 3014 shared risk links).

	Analyzed SFHR network		Full SFHR network	
	Giant	Entire	Giant	Entire
Order	199	402	1261	3265
Size	275	403	1364	3014
Min degree	1	1	1	0
Max degree	26	26	33	33
Mean degree	2.76	2.00	2.16	1.85
Diameter	15	15	16	16
Average path length	5.37	5.34	7.07	7.02
Transitivity	0.11	0.13	0.03	0.02
Cluster number	12	-	49	-

Analyzed SFHR network: Network used for causal inference in this study.
Full SFHR network: Network created from the original SFHR stud without loops or multiple edges.

In the main analysis in this study, we used a part of the original SFHR data, those with all variables need to conduct causal inference, including the information of individual attributes of 402 enrolled participants and of 403 shared risk between a pair of the participants. This means the network structure of the analyzed PWIDs network can be different with that of the original SFHR PWIDs' network that has 767 enrolled participants, 2,498 disenrolled subjects, and 3014 shared risk links. To compare how much the analyzed SFHR PWIDs network is different from the original SFHR PWIDs network, we calculated descriptive network characteristics for both networks.

Table A-2 summarizes the calculated descriptive statistics. In graph theory and network data analysis, the numbers of nodes and links are referred to as order and size, respectively. The proportions of nodes which included in the giant component in each network were similar to each other. That is, the giant component of analyzed SFHR network included about 50 % of nodes while the giant component of full SFHR network included 40 % of the entire nodes. However, there was a perceptible

difference in the proportions of links included in each network. That is, the giant component of analyzed SFHR network included about 70 % of links while the giant component of full SFHR network included 45 % of the entire links. The statistics related to degree showed similar values in each network. In graph theory, distance is defined as the shortest path between a pair of nodes in a network and diameter is defined as the longest distance in all possible pair of nodes in a network. Interestingly, the values of diameter were almost identical in analyzed and full SFHR network while the latter network has much larger in order and size. In network science, average path length is considered as a measure to know the efficiency of information transmission in a network. In the comparison of the values (i.e. average path length) of the analyzed and full network, those two values did not show substantially different. Transitivity or clustering coefficient is defined as the fraction of the number of triangles to connected triplet and is a measure of global clustering. Though there is little difference in the values of transitivity in comparing those of the giant component and entire network in each SFHR PWIDs network (i.e. analyzed and full), the values of transitivity are slightly different in the comparison between analyzed and full SFHR network. Since transitivity in full network is lower than that of the analyzed network, it can be interpreted that full SFHR PWIDs network has less dense structure compare to the analyzed network.

All in all, there is no considerable difference in structures between the analyzed SFHR PWIDs network and full SFHR PWIDs network. However, it should be kept in mind that there is still possibility that individual attributes for unenrolled subjects are substantially different with those for participants included in this study.

BIBLIOGRAPHY

Andrich D. Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement*. 1978 Oct;2(4):581-94.

Barkley BG, Hudgens MG, Clemens JD, Ali M, Emch ME. Causal Inference from Observational Studies with Clustered Interference. *arXiv preprint arXiv*:171104834. 2017.

Blanchard-Fields F, Hertzog C, Horhota M. Violate my beliefs? Then you're to blame! Belief content as an explanation for causal attribution biases. *Psychology and Aging*. 2012;27(2):324.

Curtis R, Friedman SR, Neaigus A, Jose B, Goldstein M, Ildefonso G. Street-level drug markets: Network structure and HIV risk. *Social Networks*. 1995;17(3-4):229-49.

Friedman SR, Curtis R, Neaigus A, Jose B, Des Jarlais DC. Social networks, drug injectors' lives, and HIV/AIDS: *Springer Science & Business Media*; 2006.

Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social Network Strategies to Address HIV Prevention and Treatment Continuum of Care Among At-risk and HIV-infected Substance Users: A Systematic Scoping Review. *AIDS and Behavior*. 2017;21(4):1183-207.

Grimes PW, Millea MJ, Woodruff TW. Grades—Who's to blame? Student evaluation of teaching and locus of control. *The Journal of Economic Education*. 2004;35(2):129-47.

Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology*. 1991:331-8.

Hudgens MG, Halloran ME. Toward causal inference with interference. *Journal of the American Statistical Association*. 2008;103(482):832-42.

Jacoby J, Matell MS. Three-point Likert scales are good enough. *Journal of marketing research*. 1971 Nov 1;8(4):495-500.

Kolaczyk ED, Csárdi G. Statistical analysis of network data with R: *Springer*; 2014.

Likert R, Roslow S, Murphy G. A simple and reliable method of scoring the Thurstone attitude scales. *Personnel Psychology*. 1993 Sep 1;46(3):689-90.

Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Review of Public Health*. 2000;21(1):121-45.

Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, et al. Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. *The Lancet*. 2008;372(9651):1733-45.

McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*. 2001 Aug;27(1):415-44.

Newman M. Networks: an introduction: Oxford university press; 2010.

Papadogeorgou G, Mealli F, Zigler C. Causal inference for interfering units with cluster and population level treatment allocation programs. arXiv preprint arXiv:1711.01280. 2017 Nov 3.

Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. *Biometrics*. 2014;70(3):731-41.

Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs: *General and Applied*. 1966;80(1):1.

Samejima F. A general model for free-response data. *Psychometrika Monograph Supplement*. 1972 Mar.

Saul BC, Hudgens MG. A Recipe for inferference: Start with Causal Inference. Add Interference. Mix Well with R. *Journal of Statistical Software*. 2017;82(1):1-21.

Sobel ME. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association*. 2006;101(476):1398-407.

Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. *Statistical methods in Medical Research*. 2012;21(1):55-75.

Valente TW. Network interventions. *Science*. 2012 Jul 6;337(6090):49-53.