The Role of Network Structure in HIV Transmission among Injecting Drug Users in the Philippines: A Tale of Two Cities

Nalyn Siripong

Spring 2014

Brian W. Pence, Chair

Committee Members:

Ann Dennis

Jim Moody

Kimberly Powers

Annelies van Rie

Table of Contents

1 Specific Aims **1**

2 Background and Rationale **3**

3 Significance **8**

4 Innovation **9**

5 Approach **10**

5.1 Conceptual Framework **10**

5.2 Study Design and Data Collection **11**

5.2.1 Study Setting 11

5.2.2 Study Population 12

5.2.3 Sampling Methods 12

5.2.4 HIV-1 Genome Sequencing 12

5.2.5 Preliminary Data 13

5.3 Network Measures **14**

5.3.1 Degree 14

5.3.2 Homophily 14

5.3.3 Transitivity and Clustering 16

5.3.4 Social Distance 16

5.4 Methods **16**

5.4.1 Aim 1 17

5.4.2 Aim 2 20

5.4.3 Aim 3 23

5.4.4 Statistical Power and Sample Size 25

6 Summary of Public Health Impact **27**

7 Timeline **28**

8 References **29**

List of Figures

Figure 1. HIV among injecting drug users grows rapidly from almost zero to prevalence as high as 40-50% 3

Figure 2. A plot of the $\log_{10}$ of degree (number of partners) shows that scale-free sexual networks from Sweden follows a power-law distribution 5

Figure 4. This comparison of two networks demonstrates the importance of structure on HIV infection risk. A's risk of becoming infected is higher on the highly clustered network even though A has the same number of ties in both network structures. 11

Figure 5. The phylogenetic tree of all HIV sequences from the Philippines shows predominantly AE and B subtypes, but also includes CRF01_A1, G, CRF02_AG and C clades compared to the Asia region, which has AE, B, BC and C. 14

Figure 6. Three friendship networks of 6 females (white) and 6 males (blue) compare ties chosen: (a) at random; (b) with some gender homophily; and (c) with gender homophily and higher transitivity .. 15

Figure 7. Transitivity implies that if A is friends with B and C, then B is also friends with C. 16

Figure 8. Map of HIV-1 and the p51 RT region of *pol* gene that is the target for sequencing (from: http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html) 21

Common Abbreviations

AIDS	Acquired Immunodeficiency Virus
ART	Antiretroviral therapy
CRF	Circulating recombinant forms
ERG	Exponential random graph (or ERGM, for ERG models)
HCV	Hepatitis C virus
HIV	Human Immunodeficiency Virus
IDU	Injecting drug user
ML	Maximum Likelihood
MSM	Men who have sex with men
NJ	Neighbor joining
PLHIV	People living with HIV
RDS	Respondent-driven sampling
RT	Reverse transcriptase
SI	Susceptible-infected
SIR	Susceptible-infected-recovered

1 Specific Aims

Over 11 percent of the 14 million people who inject drugs worldwide are infected with HIV [1]. Most HIV epidemics in injection drug using populations are marked by rapid spread of infection, with prevalence rates in some communities rising from 0 to 50 percent in less than a year [2]. Given the high disease burden and susceptibility to rapid epidemic outbreaks, understanding HIV transmission dynamics among injection drug using populations is critically important.

Injection epidemics are characterized by an initial rapid increase in HIV incidence followed by a plateau at some steady-state HIV prevalence. Transmission may be interrupted or facilitated through injecting behaviors like needle sharing and injecting frequency, which are the focus of most prevention efforts among injecting drug users. Yet network structure, or an injection drug user's choice of injecting partners, also strongly influences the spread of infection. Limitations in data and analytic methods have precluded use of network information to target and design more effective prevention interventions.

Two distinct HIV outbreaks among injecting drug users in two neighboring communities in the Philippines offer a unique opportunity to disentangle the relative contributions of individual behaviors and network structure on transmission dynamics. In Cebu City, HIV prevalence among people who inject drugs rapidly expanded from 0% in 2007 and 0.6% in late 2009 to 50% in 2010 and 53% in 2011. In neighboring Mandaue, meanwhile, HIV prevalence among injecting drug users remains only 3.5%. Differences in injecting behaviors do not completely explain the differences in HIV prevalence, with a majority of people sharing drugs and injecting equipment in both cities. Thus identifying differences in the network structures between the cities and formally assessing the relative contributions of individual behaviors and network structure to transmission through mathematical modeling is critical to understand the divergent epidemics in these two cities.

We hypothesize that the networks of injection drug users in Cebu City and Mandaue have different network structures, which explains the differences in HIV prevalence in the two cities.

Our specific aims are:

Aim 1: Compare network structure in Cebu City and Mandaue to identify important differences that correlate with differences in HIV prevalence.

Surveillance data were collected on a network of injecting drug users using respondent-driven sampling recruitment methods. We will use the network-related information collected in this survey to map the injecting drug user networks in the two cities. We will compare network structures using statistics on component size, homophily, social distance, and clustering. To do this, we will simulate a network from an exponential random graph (ERG) model aligned with initial network parameters, using three primary network measures: degree, homophily and clustering.

Aim 2: Examine phylogenetic clustering of HIV infections detected in Cebu City and Mandaue.

We will examine phylogenetic relationships between HIV nucleotide sequences collected from infections detected in surveillance to identify important transmission clusters and describe how they are connected between the two cities. We hypothesize that this phylogenetic analysis will identify all IDU infections as a single large cluster that is phylogenetically distinct from MSM or other risk groups, but when we examine these trees at a finer level of detail, we will find that most of the transmission is happening within each city, as opposed to across the two city boundaries.

Aim 3: Construct a mathematical transmission model to compare the relative contributions of the network and individual injecting behaviors to the epidemic dynamics in Cebu City and Mandaue.

Based on the IDU networks (Aim 1), we will reconstruct the HIV epidemics in each city using an agent-based modeling approach. We will validate the model results by comparing characteristics of infections with the phylogenetic transmission clusters identified (Aim 2). If the model can successfully replicate the

HIV dynamics in these two cities using similar biological parameters, we can be confident that our network simulations are representative of the true situation, as observed in the two cities.

If our model can successfully replicate the epidemic patterns observed, we will then use it to simulate scenarios of different behaviors or network configurations to evaluate their impact on disease transmission. In addition, we will use the model to identify those network measures that are most predictive of disease risk. This knowledge will enable more effective and efficient HIV prevention strategies that target individuals with the most important positions in the IDU network with respect to HIV transmission.

A better understanding of how an individual's behaviors and the underlying network structure interact to promote or prevent HIV transmission will enable the design of more targeted HIV prevention programs for injecting drug users that are more effective and efficient at slowing epidemic growth and expansion.

2 Background and Rationale

2.1 Illicit injecting drug use is an important issue and a growing global public health concern.

According to the latest estimates, between 167 and 315 million people use illicit drugs and 14 million of them administer drugs by injection. Use of illicit drugs has serious implications not only for health but also for the economy of affected countries. The impact of illicit drug use is estimated at almost $200 billion in the United States alone [3].

2.2 HIV epidemics among people who inject drugs are a serious public health concern.

Among the estimated 14 million injecting drug users (IDUs) worldwide, 1.6 million (11%) are infected with HIV [1]. HIV infection presents a serious disease burden among IDU populations, exceeding 40% prevalence in countries across Southeast Asia, including Thailand and Indonesia [4-6]. In these concentrated epidemic settings, HIV epidemics among IDUs precipitate larger-scale heterosexual epidemics [7-10], which further underscores the urgent need to find effective ways to reach and deliver prevention services to injecting drug users.

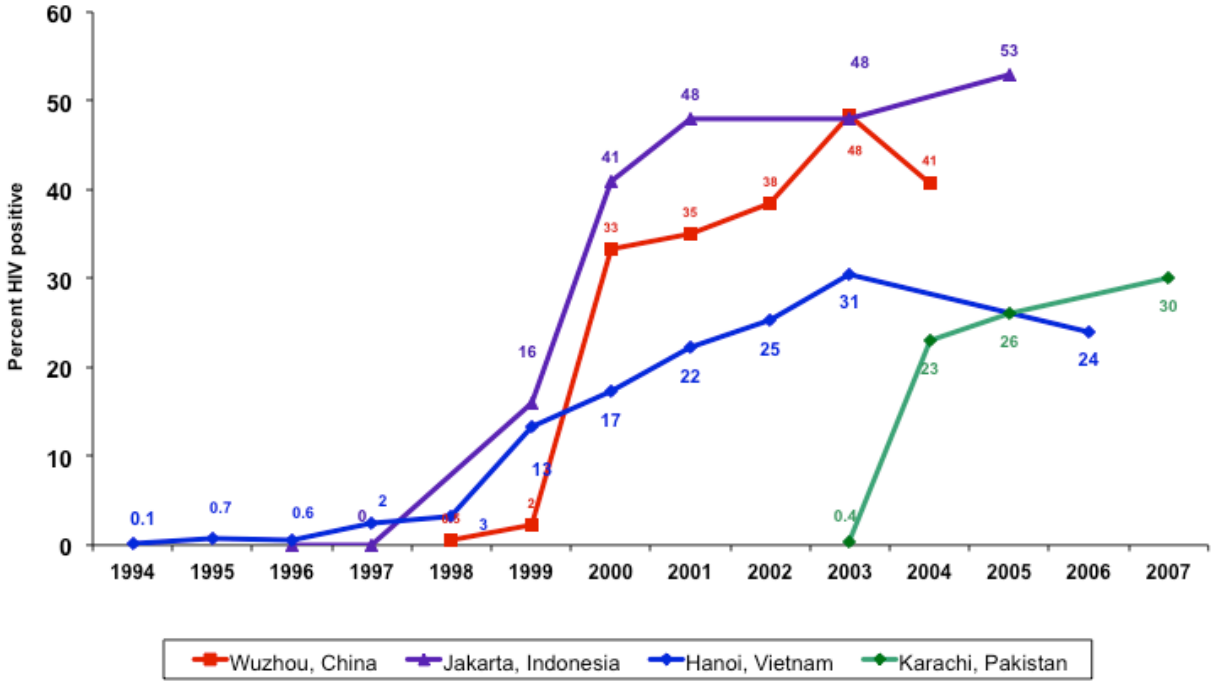


Figure 1. HIV among injecting drug users grows rapidly from almost zero to prevalence as high as 40-50% **[11]**.

A striking and unique feature of IDU epidemics is the rapid HIV growth, from virtually 0% prevalence to a steady-state level of 20-50% HIV prevalence within 3-6 years (see Figure 1). From Southeast Asia to China, England to Greece, and even in parts of the US, this transition has been characterized by an initial period of very high HIV incidence, with rates reaching as high as 12.5 cases per 100 person-years [10]. As an increasing number of injecting drug users became infected, the pool of uninfected and susceptible IDUs remaining was depleted. As the pool of susceptible IDUs decreased, the incidence rate fell to rates as low as 2.4 cases per 100 person years [12] and settled at a rate equal to the turnover rate in the injecting drug user population.

2.3 The rapid HIV spread among injecting drug users results from biological and behavioral factors.

At the biological level, HIV transmission through needle sharing is widely recognized as an efficient mode of transmission but estimates of actual per-exposure transmission rates are not well documented. Available studies of the probability of transmission through injecting drug use range from 6 to 200 transmission events per 1000 contacts [13], which far exceeds the rate of male-to-female and female-to-male transmission in heterosexual (vaginal) sex, estimated at less than 1 per 1000 exposures [14].

Beyond the heightened transmission rates, IDUs often engage in risky injecting behaviors that increase opportunities for HIV exposure. In populations where HIV is circulating, IDUs who do not regularly report clean needle use are at considerably higher risk of infection (HR: 5.5) and injectors who share "almost all the time" have an increased odds (OR: 13.2) of HIV compared to those who never share [15,16]. Other practices, such as frequent injecting [17], and syringe-mediated sharing of drugs through front- or back-loading [18,19] also create opportunities for exposure and infection.

2.4 Risky behaviors alone, measured at the individual level, may be insufficient to capture risk of HIV infection; network structure can create important bottlenecks that will slow transmission.

After the introduction of HIV into a community, high-risk exposure through unsafe injection practice is necessary but not sufficient to facilitate the "explosive" HIV epidemics typically observed among injecting drug user populations. In some communities, HIV prevalence among IDUs may remain unusually low, despite their risky injecting behaviors [20]. This unusual finding may be explained by the network structure. If infection is introduced in a small, closed groups of IDUs who share with no one but each other, it is unlikely that we will observe the typical rapid disease growth described in section 2.2. Instead, most of the members within one small group will become infected over time, but if they do not leave to inject outside the group, then HIV prevalence will remain quite low, as observed [20].

In other words: even more than the *act of* sharing, it is *with whom* one shares that confers HIV risk. And often, it is not simply one's immediate partner but one's partner's partners who may influence that individual. To fully understand the interaction of connections between people and how they influence HIV risk, we must consider networks at two levels: the immediate network surrounding any given individual (the ego-network) and the overall network (the global or sociometric network), which describes the distribution of different types of people and how they are connected to each other.

2.5 Network analysis provides important insights into understanding behavioral patterns and disease transmission dynamics.

Networks consider how an individual (*ego*) is connected to others (*alters*) through specific *ties*, *edges,* or *paths;* and network analyses consider both the overall network structure and the individuals within it. Network features such as membership to certain types of groups or occupation of particular positions in the network [21-24] may increase or reduce risk of infection for different people. For example, IDUs in New York City [23] who were members of a 2-core (a group of 4 or more people, where everyone is no more than two path lengths from every other person) were more likely to engage in a number of risky behaviors and were also more likely to be HIV-infected, compared to those who were not in a 2-core. Another study of sexual networks found that people connected to a network larger than two were at much higher risk of HIV infection [22]. In other words, people in mutually monogamous relationships were strongly protected from infection, compared to everyone else with any partner – even after adjusting for other risk factors. Even a person who had only one partner was at increased risk if his or her partner was not monogamous. This was one of the first studies to collect and incorporate data on behaviors, networks and STI infection over time to confirm what appears to be a relatively intuitive finding.

2.6 Network measures can elucidate important aspects of HIV risk and transmission.

A number of quantitative measures are used to describe networks. Degree centrality, for example, is a measure of connectedness at the individual level. It is defined as the number of ties an individual has to other people in the network. In an HIV injecting network, an IDU's degree is the number of people with whom he shares injections. A person with high degree centrality is well connected and thus has a higher potential for HIV exposure. Increasing degree is highly correlated with higher risk of HIV infection. If a pair of HIV-negative IDUs injects exclusively with each other, then they both have degree one, and both are at very low risk of infection. Their network configuration would be analogous to the mutually monogamous heterosexual partnership described above. As soon as one person in that pair begins to inject with others, the risk of infection for *both* IDUs increases.

The frequency distribution of individuals' degree centrality is a descriptive summary of the network overall and may provide insight into its structure. For example, social networks are often described as following a "small world" concept, in which most people are connected to only a few others, but a few people connect to everyone and thus create this "small world" [26]. Sexual networks are one type of "small world" network that follows a scale-free form, which can be visually tested by looking at the log-log frequency plots (Figure 2). Scale-free networks follow a power-law distribution, which means that the log-log frequency plot follows a straight line, as shown in the data on sexual networks in Sweden [25]. However, the term "small worlds" can also apply to other frequency distributions as well [27].

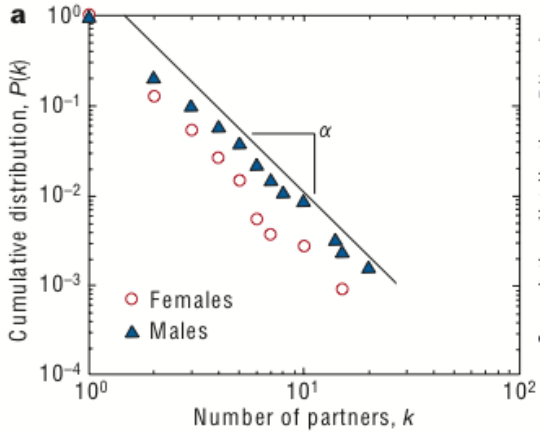


Figure 2. A plot of the $\log_{10}$ of degree (number of partners) shows that scale-free sexual networks from Sweden follows a power-law distribution [25]

A number of additional network measures describe individuals' roles in spreading or perpetuating transmission chains in a network. Many of these measures require a complete census of the full network. Our recruitment methods did not collect complete network information to maintain the anonymity of all participants and their partners. There are several additional measures of centrality that may explain transmission dynamics, but available studies suggest that most of these measures are very highly correlated across a range of networks [28,29].

2.7 The composition of a person's network can influence potential risky behaviors.

People make choices about their ties, whether those ties are social, sexual or injecting. Individuals often tend to connect with others who are either very similar (in that they share similar interests or have common values) or decidedly different (e.g., a heterosexual network would always choose people of the opposite sex) from themselves. Homophily refers to the tendency to befriend or connect with people who have similar characteristics; it measures the degree of assortative mixing on some characteristic or preference among people in a network. For example, in parts of China, injecting drug users tend to form groups with people in the same ethnic group and accordingly exhibit different injecting norms within their networks [30]. In Australia, IDUs segregate by ethnic background (Vietnamese origin vs. non-Vietnamese) [31].

By the same logic, heterophily occurs when people are connected to others who have decidedly different or opposite characteristics. For example, the presence of an older IDU in one's network (which may be described as age heterophily) is associated with increased risky practices such as renting syringes or visiting shooting galleries [32,33]. Preferences for homophily or heterophily along particular characteristics are important social processes that determine network structure, and this underlying structure can, in turn, impact risky behaviors.

2.8 Molecular analysis of genetic sequencing of HIV is an important tool for understanding HIV transmission dynamics at the population level.

Even in the early years of the HIV epidemic, genetic subtyping was used to define the spatial boundaries of viral subtypes and to the trace the spread of disease [34]. For example, a description of the spread of different subtypes by geographic region was used to demonstrate that HIV probably spread through injecting drug users, and transmission patterns followed different drug trafficking routes in Southeast Asia [35]. Phylogenetic studies also described patterns in the spread of HIV strains across countries and through different risk groups [36-39]. One study of the spread of circulating recombinant form CRF01_AE in Vietnam traced it from Thailand, through the Vietnamese heterosexual population and finally to injecting drug users [40].

Phylogenetic analyses incorporate demographic and clinical information, which supports further epidemiologic investigation of disease spread. One analysis of publicly available HIV-1 nucleotide sequences in the United Kingdom identified several independent transmission clusters sparsely spread over space and time [41,42]. This pattern suggests that HIV spread is not driven by core groups but follows a pattern of more sporadic outbreaks. Researchers used phylogenetic analysis to show that MSM in Mississippi do not choose sexual partners according to geographic proximity. Instead, phylogenetic analysis of infection patterns indicate that they tend to seek partners with similar demographic characteristics - likely a result of greater homophily by age and race in this sexual network [43,44]. The addition of behavioral and geographic data to their analysis provided a much richer understanding of how HIV is spread in the MSM population.

2.9 Mathematical modeling is a useful tool for testing hypotheses about underlying mechanisms by simulation.

Mathematical (transmission) modeling is an important epidemiologic tool that simulates the effects of biological and behavioral factors on disease transmission in a population. Unlike traditional statistical regression models, mathematical models usually require that mechanisms of transmission are explicitly defined. These models can simultaneously accommodate multi-factor and multi-step mechanisms of disease. Models that have been validated may further be used to compare scenarios of potential outcomes under different circumstances or environments.

Mathematical models are often used to estimate the impacts of past prevention programs or to compare different policy scenarios in a particular population or region. The Asian Epidemic Model was used to estimate the impacts of the success of the “100% Condom Use Program”, which slowed the growth of the HIV epidemic in Thailand in the 1990s. The investigators estimated that for every $1 invested in this prevention strategy, the Thai government saved $43 in averted treatment costs [45]. In a study of the SHAKTI (Stopping HIV/AIDS through Knowledge and Training Initiative) program in Bangladesh, the IDU 2.4 model was used to estimate that this prevention effort reduced HIV incidence by 90%. Their models predicted that without the SHAKTI program, HIV prevalence in Dhaka would have risen as high as 42% over 8 years, in contrast with the 10% observed in surveillance at the time [46]. More recently, mathematical models have also been used to prospectively estimate the impacts of a policy of providing widespread antiretroviral treatment as a prevention strategy [47,48]. All these models have applied a compartmental approach, which assumes that groups or sub-groups of people meet and interact at random. At an aggregate level and over many years, this approach may be appropriate to estimate the process of HIV transmission and to compare the impact of different prevention program scenarios. However, such models cannot capture the importance of heterogeneity of individuals within a network, and would not reflect the features of the social and injecting networks.

2.10 Agent-based modeling allows us to model individuals or groups in greater detail.

Because networks create structures that characterize individuals or groups in very different ways, agent-based models are likely a more realistic approach to modeling the diffusion of infection across a network.

Agent-based models have been used to simulate the potential impacts of concurrency on HIV transmission in a heterosexual network [49], to compare the effects of different combination HIV prevention strategies for injecting drug users [50], or to estimate mean time to Hepatitis C infection among IDU in Australia [51]. The application of agent-based models is more tractable as a result of rapid increases in computing power in the last decade.

2.11 Summary

Injecting drug use is a dangerous practice with serious public health consequences, especially the heightened risk of HIV infection through shared use of injecting equipment. Because transmission is efficient and IDUs tend to inject quite frequently, the introduction of a single HIV infection is typically followed by rapid spread, with HIV prevalence jumping from 0% to 50% in as little as 6 months [2]. However, in some communities, the spread of HIV remains limited, despite highly prevalent risky injecting behaviors. Individual behaviors may be comparable between those in high HIV-prevalence communities and those in low HIV-prevalence ones. Understanding the network structure that describes the dynamics of HIV transmission may help elucidate how rapidly and to what extent HIV will spread across a network.

In a brief review of the literature on three analytic methods relevant to HIV transmission among injecting drug users (IDUs), we found no other study to use an RDS-sample of IDUs to reconstruct networks and simulate disease transmission across them. Studies of social networks among injecting drug users are challenging because of difficulties in collecting complete data; our study proposes the application of newer methods that overcome this difficulty by simulating network structure from partial data.

By incorporating network and behavioral data into phylogenetic analysis, we can provide more context and strengthen the interpretability of our results. Mathematical modeling simulates HIV transmission dynamics, based on what we have learned about the network structure and potentially linked infections. Triangulation of all these methods to describe development of the HIV epidemic may offer greater reliability and credibility of the results.

3 Significance

This work has important implications for surveillance and prevention of HIV risk among IDU populations.

3.1 Social networks and risk networks are recognized as important facilitators of infection, but they are still understudied in infectious disease epidemiology.

Our study considers how different network structures may explain differences in observed HIV prevalence in two cities in the Philippines. Similarities in the local cultures should improve comparability of the two networks. In addition, we will model disease transmission across each network to demonstrate the importance of network structure in explaining disease transmission processes. This work will help us identify factors that explain the unusual disparity in HIV prevalence. Unraveling the linkages between and across populations will help identify transmission drivers and important targets for prevention [33,52,53]. Elucidating how individual behaviors and network structures jointly facilitate or impede HIV transmission may offer important insights on early warning signs or identify specific individuals whose positions in the network may offer key opportunities for prevention.

3.2 This study leverages the use of existing and ongoing data collection mechanisms.

Our analysis uses data from the Philippine Integrated HIV Behavioral and Serological Surveillance (IHBSS) survey. These surveys do not follow individuals over time, but if each sample is representative of the population, then we should be able to use these data to describe demographic and behavioral changes in the overall at-risk population. For example, comparable surveillance data have been used to reflect increased condom use that coincided with increased coverage of HIV prevention services in a population of female sex workers in Nepal [54].

Moreover, the Philippine IHBSS survey is repeated every 2 years to monitor for "hot spots" of risk behaviors and potential disease spread. If we find survey questions that identify individuals at particularly high risk of acquiring or transmitting infection, we can validate these findings by repeating the network analysis in the sample from the next round of surveillance.

3.3 Identifying individuals who occupy important network positions can guide prevention and treatment efforts.

If we can locate and characterize those IDUs who are at particularly high risk of infection and who are important for continuing or expanding HIV transmission, then these methods have broader implications for identifying targets for prevention, testing and treatment. We know that in a sexual network, being in a mutually monogamous relationship is protective of disease risk, compared to having multiple partners or just one partner who is not monogamous [22]. Certain types of group membership or positions in a network were also found to associated with increased HIV risk in IDU networks [23]. These studies showed the importance of networks in facilitating risk and infection, but they required intensive data collection to enumerate the whole network. If we can extract such information with partial network data, this work could have powerful implications for the utility of network analysis for public health purposes.

3.4 This work could serve as a prototype for the application of ERG models to RDS data.

Respondent-driven sampling methods have been used for HIV surveillance among high-risk populations in over 100 countries throughout the world [55]. If we can demonstrate the feasibility of using ERG models for network simulation and their validity in identifying targets for prevention and testing, ERG models could become a powerful tool for analyzing RDS data to inform public health action.

4 Innovation

We capitalize on the unique opportunity to identify key differences in network structures that may explain differences in HIV transmission dynamics between two adjacent cities. The innovation of this study lies in the synthesis of several forms of data collected from the same two networks. Using data from the ongoing Integrated HIV Behavioral and Serological Survey (IHBSS), we employ different types of data and methods to describe the network structures and to reconstruct the HIV epidemics among IDUs in two cities (Cebu City and Mandaue) in Greater Metro Cebu. As this surveillance continues, we can evaluate the validity of our approach by replicating the analysis in these IDU populations over time.

4.1 Our analysis will extend beyond studies of injecting behaviors to measure the effects of risk networks on individual and population-level HIV transmission dynamics.

Although networks are recognized as a facilitator of rapid HIV spread, they are logistically difficult to study due to missing data (on people or ties between them) and concerns about protecting confidentiality. We use information and network statistics extracted from data on a partial network to infer the complete network structure in the total IDU populations in each city. Advances in computing power and the identification of new constraints that reduce parameter estimation space have made these models more viable [56,57].

4.2 We link sequencing data with information about network position and risk behaviors to provide a more complete picture of transmission dynamics in the two cities.

Phylogenetic analysis of HIV has been used in a number of ways to advance epidemiologic research. Analyses have identified linked transmission clusters of infection at the inclusion or exclusion of specific individuals [58,59]. Historical studies have used phylogenetic analyses to trace the spread of specific HIV subtypes or circulating recombinant forms (CRFs) across geographic borders, population sub-groups, and modes of transmission [34,40,41,60,61], but very few have also had access to behavioral and demographic data. Studies where this information was linked [41-43,62] presented important insights about the geographic and social spread of disease; however, all of them focused primarily on sexual transmission of infection. This study will combine sequencing data with network data as well as demographics and injecting practices to characterize the transmission clusters identified and identify those network positions and behaviors that may signal heightened risk of HIV infection.

4.3 Mathematical modeling and simulation will allow us to explore the relative importance of behaviors and network structure in HIV transmission at the population and individual level.

Synthesizing all available data into a mathematical transmission model will provide a complete picture of the important interactions between individual behaviors and network structure. Epidemiological studies of networks can be challenging because of difficulties in identifying and interpreting different network measures. Studies of HIV and HCV transmission among injecting networks discuss the importance of membership to particular groups structures [23], but it is difficult to translate these results into public health action. Using our models, we can explicitly hypothesize and simulate the mechanisms through which the network structures may influence disease spread. We could also use the model to pinpoint an index or other measure that could be captured in routine surveillance and describe network structure.

This will be one of the first studies to reconstruct a complete network from respondent-driven sampling data and simulate disease transmission along the network to validate our network simulation. These analyses may provide unique information to target, design and deliver prevention interventions to minimize the impact of the HIV epidemic.

5 Approach

To describe the effects of network structures on HIV growth in Cebu City and Mandaue, we will construct a mathematical model that will compare the transmission of HIV in these two populations. This work comprises three parts: construction and comparison of each city's network based on surveillance data (Aim 1); evaluating genetic and spatial distribution of observed infections based on sequencing data (Aim 2); and modeling disease transmission across each network (Aim 3). By reconstructing each IDU network and simulating HIV transmission across them, we hope to better understand the importance and impact of network structure in facilitating the spread and growth of this infectious disease in this population.

5.1 Conceptual Framework

Two minimal conditions are necessary for HIV transmission to occur as a result of injecting drug use: (1) one IDU must be infected with replicating virus; and (2) the HIV-infected IDU must use injecting equipment (cookers, needles, or drugs) before sharing it with his uninfected partner. Network structures play an important role in creating these two conditions, as laid out in Figure 3. This diagram has four arrows, which we discuss in greater detail here.

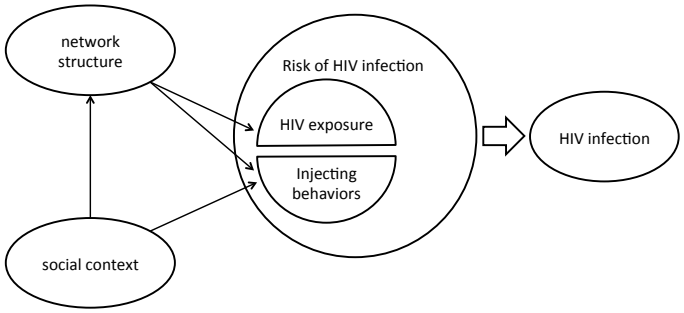


Figure 3. Conceptual diagram showing pathways that describe how social context and networks influence HIV risk and subsequent infection.

5.1.1 Network structure → HIV Exposure

Network structures are particularly important because they capture the effects of social ties beyond the direct interactions between individuals. The importance of network structures is illustrated in Figure 4, which is a comparison of two simple networks. Each of the networks has 9 nodes and 10 ties, so that the network densities are equal. If we consider just individual A, or even the pair A-B, both persons A and B appear to have the same level of risk, as measured by their ties. The surrounding network structures, however, imply very different contexts and risk profiles – especially for person A. If we assume that person F was HIV-infected, then on the first network (left), A and B could simultaneously acquire infection through different partners (if F infected both G and C, who subsequently infected A and B, respectively). Under a slightly altered network structure (right), person A's risk of infection is only through person B.

5.1.2 Network structure → Injecting Behaviors

The overall network structure can also influence a person's HIV risk by changing his injecting behaviors. The formation of networks arises from social processes such as assortative mixing (the tendency for people to connect with people of similar characteristics or interests) or transitivity (my friends are also

friends with each other).[1] These network properties also have an effect on injecting practices [63], which makes the study of these structures of increasing interest and importance in public health.

5.1.3 Social context → injecting risk

Social context and environment are important determinants of the level and frequency of injecting risk. In societies with a heavy police presence, IDUs are more inclined to borrow used syringes. Needles are often left in public spaces (e.g., public restrooms) for common use, so that people may avoid the risk of arrest for carrying a needle [64]. In shooting galleries, IDUs will also rent needles for similar reasons of avoiding arrest [65].

5.1.4 Social context → network structure

The social environment also drives the formation of network ties, which serve as the entry points for HIV exposure. For example, IDUs with lower incomes may pool their money to purchase and share drugs, which creates a drug-sharing network as a vehicle for spreading HIV infection. Here, poverty in the population generates the formation of networks; and the networks, in turn, increase the potential for HIV exposure. These network structures and parameters will be discussed in greater detail in our methods.

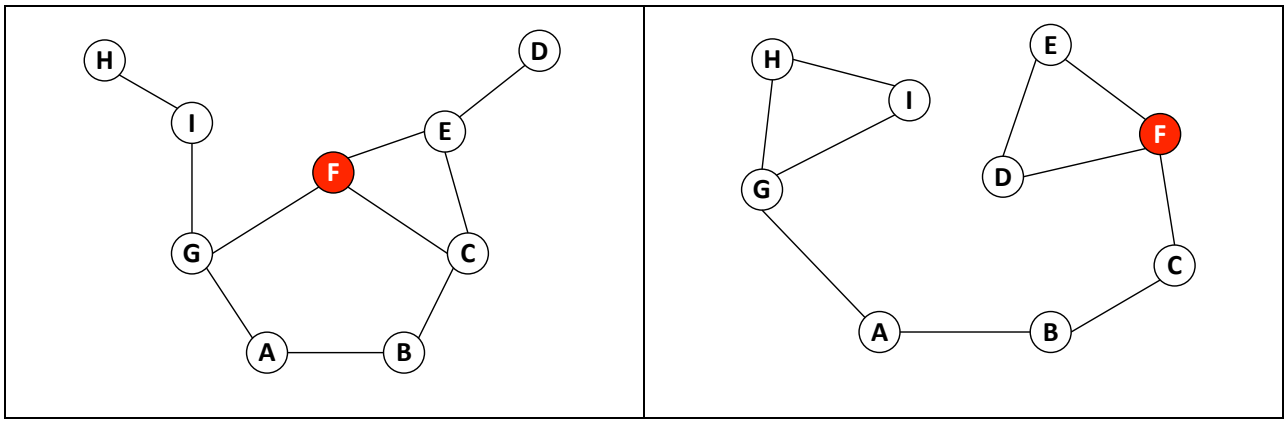


Figure 4. This comparison of two networks demonstrates the importance of structure on HIV infection risk. A's risk of becoming infected is higher on the highly clustered network (right) even though A has the same number of ties in both network structures.

5.2 Study Design and Data Collection

We will analyze existing data that was collected as part of the IHBSS surveillance in the Philippines. These surveys are used to identify areas of high risk and high disease burden among key affected populations (female sex workers, injecting drug users, and men who have sex with men). Surveys collect information about demographic and behavioral characteristics, as well as biological specimens for HIV, HCV and syphilis testing and sequencing. The National Epidemiology Center in the Department of Health has maintained and expanded this active surveillance system since 2005. It now covers around 20 cities throughout the Philippines, including Cebu City.

In response to the HIV epidemic observed in Cebu City in 2009, surveillance was expanded to neighboring Mandaue in 2011. Our analysis will focus on data collected in 2013 but we will incorporate earlier data for model parameterization and validation in Aim 3.

5.2.1 Study Setting

Metropolitan Cebu is a region that encompasses seven cities and six municipalities. Home to almost 2.5 million people, it is the second largest city in the Philippines. Almost half of the estimated 6,000 injecting

[1] There are many other social processes (e.g., balance) and phenomena (e.g., the importance of weak ties) that are important; here we focus on the two most relevant to our research question.

drug users in Cebu province are located in Cebu City; the remaining 3,000 are spread across four cities in Metro Cebu: Mandaue, Danao, Lapu-Lapu and Talisay. There are very few distinct boundaries or borders across these cities, with the exception of Lapu-Lapu, which is located on Mactan Island.

5.2.2 Study Population

Our data come from the 2013 IHBSS surveillance round, which was conducted in Cebu City from May through June and Mandaue from June through July. Seven 'seed IDUs' (see description below) from Cebu City recruited 450 other injecting drug users into the survey, for a total of 457 people. Another seven 'seed IDUs' from Mandaue recruited 303 other eligible injecting drug users, for a total of 310 IDUs in this survey.

Injecting drug use is not high at the national level in the Philippines, but it is common practice in Cebu. Most IDUs in Cebu inject Nubain (nalbuphine hydrochloride), a narcotic analgesic often prescribed in medical care settings for expectant mothers during labor. Pooling money to purchase drugs is a common practice that facilitates drug-sharing routes. Shooting galleries are another source of infection, as injectors who go there to rent needles are unaware of the potential risk involved.

5.2.3 Sampling Methods

Injecting drug users are recruited using a modified snowball sampling (also called link-tracing) technique called respondent-driven sampling (RDS). This method samples on a network, and therefore requires specific calculations for weights and variance of population characteristics [66]. It has been used for the recruitment of hidden or hard-to-reach populations, including injecting drug users, in over 100 countries worldwide [55].

Several IDUs are selected as the "seeds", who begin the recruitment process. Seed IDUs are often selected to represent a wide array of demographic characteristics, to promote sampling from different positions and sub-groups within the network. After they complete the behavioral interview and a blood draw (testing for syphilis, HCV and HIV), these IDUs are presented with 3 coupons[2], each of which has a unique respondent ID number. They are instructed to give these coupons to other IDUs they know who are eligible for the survey. IDUs who receive the recruitment coupons go to the surveillance site to enroll. After they complete the interview and provide a blood sample, these IDUs are given 3 coupons to recruit more people into the study.

To promote recruitment, a primary incentive is provided to IDUs who complete the questionnaire and provide a blood sample. IDUs who successfully recruit others into the study (as tracked by their coupons) are also provided a secondary incentive for each additional eligible IDU recruited. This incentive structure promotes greater participation, which is why this method is particularly useful for sampling hard-to-reach populations. However, because participants are not sampled at random, we cannot assume observations are independent. In Section 5.4.1, we describe the weighting structure that is required to estimate unbiased parameters for the networked population.

5.2.4 HIV-1 Genome Sequencing

Blood samples that test positive for HIV are submitted to the Hawaii Center for AIDS (HCFA) for sequencing. RNA extraction and stabilization are performed at the STD-AIDS Cooperative Central Laboratory in Manila to prepare for shipment to Hawaii. The laboratory at HCFA amplifies and purifies the sample before analyzing nucleotide sequences for each patient. An optimal primer set, which has been identified based on its effectiveness in amplifying the target region (p51 region of the *pol* gene), with wide coverage of different subtypes and CRFs, is used for amplification. Gel electrophoresis is conducted to confirm successful amplification of PCR product. The amplified product is purified and sequenced. To

[2] In the 2013 round for Cebu City, IDUs received only 2 coupons per person, but usually (i.e., in past years and all other RDS IHBSS samples in the Philippines) each respondent is provided 3 coupons.

date, 22% (22/86 in Cebu, 2/21 in Mandaue) of samples from IDU populations in Mandaue and Cebu have been successfully amplified. However, a number of samples are still in queue for amplification.

5.2.5 Preliminary Data

The spread of HIV among IDU in Greater Metro Cebu was very rapid, comparable to other IDU settings [2,10,30,67]. As late as 2009, HIV prevalence was still less than 1% among IDUs (2/305) in Cebu City, but high rates of injecting risk raised concern. Modeling and general expert opinion predicted rapid expansion of infection in this population in the next 1-2 years. This prediction prompted an interim rapid study of the epidemiologic situation in Greater Metropolitan Cebu in 2010. In this rapid study, HIV prevalence was 56% HIV in Cebu City, and limited data about IDU from Mandaue indicated potential spread to this neighboring city.

Table 1. Risk behaviors in the Cebu City and Mandaue in 2011 are comparable but HIV prevalence appears very different

	Cebu City (N=305)	Mandaue (N=311)
HIV Prevalence	52.9%	3.5%
Visits Shooting Galleries	90.7%	74.1%
Sharing needles at last injection	71.2%	64.2%
Injected with other IDUs	70.6%	91.6%
Pooled funds to buy drugs	90.2%	92.9%

To explore the extent of HIV infection in Mandaue, this city was added as a new surveillance site in 2011. Those results indicated that IDUs in Mandaue were slightly older and less educated than those in Cebu City, but they also exhibited risky injecting behaviors — with 90% pooling funds with friends to buy drugs and a majority sharing at their last injection in both cities (Table 1). In contrast, HIV prevalence was very different in the two cities, with 53% prevalence in Cebu City and 3.5% in Mandaue.

Phylogenetic analysis was conducted on newly reported HIV infections in the Philippines through 2011, to describe transmission patterns across risk population, geographic space, and time. The most recent analysis covered 185 infections from a range of risk groups (injecting drug users, men who have sex with men, and male visa applicants) collected through surveillance and case registry data [68]. In this study, most IDU infections were subtype B (95%, 61/68), with an average genetic distance of 0.010 (±0.001).[3] This was much shorter than the genetic distance between MSM infections, which ranged from 0.023 (±0.002) among the 21 MSM with CRF01_AE, to a distance of 0.044 (±0.004) in the cluster of 10 MSM with subtype B. These very small genetic distances in IDUs suggest that most of the HIV epidemic was very recent and rapid, while HIV infections among MSM were likely introduced through different sources and over longer periods of time.

Unlike much of Southeast Asia, HIV infections in the Philippines cover a wide range of subtypes and CRFs. A phylogenetic tree, based on a BLAST search of all HIV nucleotide sequences of the *pol* gene from infections in the Philippines (Figure 5), found at least six different subtypes and CRFs in the Philippines. As described above, subtype B and the CRF01_AE are most common, but other subtypes (C and G) and recombinant forms (CRF02_AG) are also present. This wide range of clades may be a reflection of the HIV infections among returning overseas Filipino workers, who constituted almost one-third of newly reported infections in the Philippines through 2009. Since many different HIV subtypes are present in the Philippines, finding similar subtypes among IDUs in the two cities would further substantiate the hypothesis that IDUs in Mandaue are acquiring infection through sharing injections with IDUs in Cebu.

[3] Of the remaining 7 infections among IDUs, 6 were CRF01_AE and 1 was subtype C.

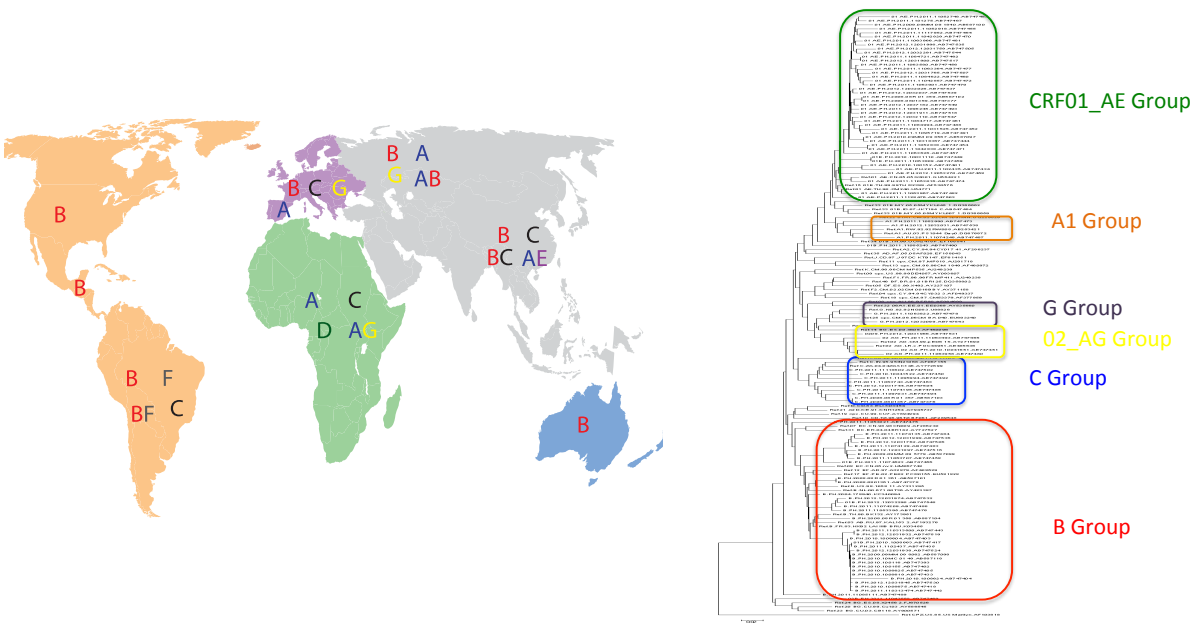


Figure 5. The phylogenetic tree of all HIV sequences from the Philippines (right) shows predominantly AE and B subtypes, but also includes CRF01_A1, G, CRF02_AG and C clades compared to the Asia region, which has AE, B, BC and C (Source: [69], *Hawaii Center for AIDS*).

5.3 Network Measures

We will construct and compare our networks using four measures: (1) degree distribution, which depicts the density of ties; (2) homophily, which measures the degree of assortative mixing; (3) social distance, which counts the number of intermediates between any two people in the network; and (4) transitivity, which describes the extent of triad closure in the network. While each of these measures can be used to ascribe network characteristics to an individual, this work will focus on how they describe the topography of the overall network. We illustrate each of these on a small hypothetical example comparing three friendship networks of 6 males and 6 females as shown in Figure 6.

5.3.1 Degree

The most basic measure of how much a network influences a person is how many ties a person has, which is also called their degree. At the network level, degree can be summarized using median or mean, but the variation is captured by the degree distribution (see Section 2.6 for a brief discussion). This is simply the frequency distribution capturing the number of people at each level of degree. The degree distribution is used as a starting point for our model simulation.

In Figure 6, all three networks have the same average degree (or average number of contacts per person), which is also reflected in the fact that network density (the proportion of all possible ties that are observed) is equal across all three pictures; but their distribution differs. In the first network, everyone has exactly three ties but in the second two networks males (blue) have two ties each and the females (white) have four.

5.3.2 Homophily

Homophily is often described as the "birds of a feather" phenomenon because it measures the extent of assortative mixing within a network. On friendship networks, people tend to choose friends with similar interests or characteristics, which drives this effect. At the individual level, homophily is characterized as the proportion of total ties (his degree) that occur among people with the same property. A girl who has

only other female friends would have a homophily statistic of 1, which indicates complete gender homophily. Similarly, a boy with exactly one male friend also has a homophily statistic of 1.

We illustrate effects of homophily in Figure 6b. In this network, the average degree and network density are still the same, but females (white) have 4 ties each and males (blue) have 2 ties. Homophily is evident on this network, where a greater than expected number of ties is between people of the same gender: half of all ties are between girls, one-sixth are between boys, and the remaining one-third are mixed-gender pairs.[4]

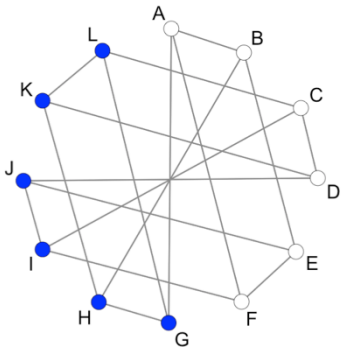
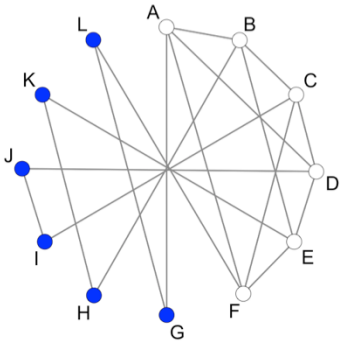
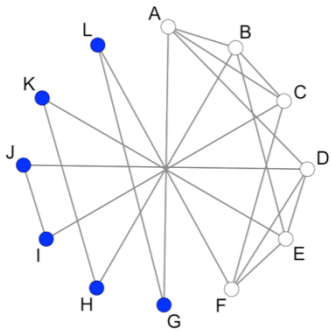
(a) Random network	(b) Gender homophily	(c) Homophily with transitivity
		
Network density = 0.27 Average degree = 3 §Mean distance (n) = 2 *Mean distance (A) = 2 Homophily (A) = 0.67 Transitivity (A) = 0	Network density = 0.27 Average degree = 3 §Mean distance (n) = 2 *Mean distance (A) = 1.67 Homophily (A) = 0.75 Transitivity (A) = 0	Network density = 0.27 Average degree = 3 §Mean distance (n) = 1.91 *Mean distance (A) = 1.58 Homophily (A) = 0.75 Transitivity (A) = 0.17

Figure 6. Three friendship networks of 6 females (white) and 6 males (blue) compare ties chosen: (a) at random; (b) with some gender homophily; and (c) with gender homophily and higher transitivity. Note that the mean distance across all n nodes on the network (§) may not change from (a) to (b) but it does change with the additional clustering in (c); while the mean distance for a single node A (*) changes across all three networks. Note also that while the first three measures reflect total network structure, the last three measures are ascribed to nodes (see note 4 for discussion of measuring homophily at the network level.

[4] There are many different ways to measure homophily. Here, we treat homophily as a measure of assortative mixing above and beyond what is expected at random, *assuming that each person's degree remains constant*. Notice that in this example, where girls have twice the number of ties compared to boys, we would expect one-ninth (11%) of ties between two boys, four-ninths (44%) to be among girls, and another four-ninths (44%) to be mixed-gender. Therefore, our results of one-sixth of ties (16%) between two males and half of ties (50%) between two females clearly indicate assortative mixing by gender.

5.3.3 Transitivity and Clustering

Beyond the effects of homophily, transitivity is another important property of networks. In its simplest form, transitivity is described as the idea that if two ties exist across three people (see Figure 7), then the third tie (dashed line) is more likely to also occur, thus closing the triangle. This may arise from shared contact with the intermediate (person A in Figure 7) or may simply be a result of shared interests among the three.

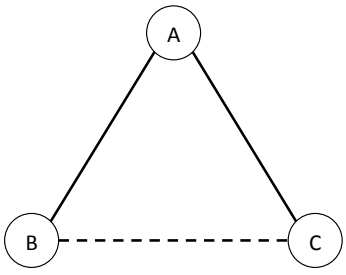


Figure 7. Transitivity implies that if A is friends with B and C, then B is also friends with C.

We can see the added effects of transitivity by comparing the second two networks, Figure 6b and Figure 6c. These two networks have the same degree and degree distribution, and the same level of homophily. However, the network in Figure 6c has greater transitivity, as evident by the presence of closed triangles, A-B-C and D-E-F.

Sometimes, transitivity may occur as an artifact of highly assortative mixing. Notice that the assortative mixing in Figure 6b has no evident transitivity (no closed triangles among 3 people of the same sex). If mixing were even more assortative on this network (for example, if females chose *only* female friends in the network), then we would observe higher transitivity, simply as a result of the heightened homophily.

The term transitivity is applicable only when considering triads or three individuals. As network size and the number of ties increases, large groups with high transitivity are described as clustered; thus, we often use the term clustering to capture both highly transitive and well connected groups.

5.3.4 Social Distance

Social distance refers to the number of ties between any two people on a connected network. Starting from an index person on the network, the distance refers to the shortest path length needed to reach any other person on the network. For example, in the network in Figure 6a, every individual has a distance of 4. This means that starting from any person (A – H) on our network, we can reach any person in 4 steps or less. To go from person A to person D in this network, for example requires four steps[5]. We can also estimate mean distance, which refers to the average of all distances from an index person to *every* other person on the network. In Figure 6a, everyone has a mean distance of 2.

When homophily and transitivity are present on the network, it is easier for some people to reach (and be reached) than others. For example, homophily among females (Figure 6b) results in a shorter distance: any female on this network (labeled A – F) can reach any other person (male or female) within 3 steps. For males, it still takes 4 steps to reach some of the people. In Figure 6b, A and D are now directly connected, reducing their distance from four (in Figure 6a) to one. Similarly, in the network in Figure 6c, several people are slightly closer to each other: to get from person A to person C required 3 steps in Figure 6a (A-G-L-C), 2 steps in Figure 6b and only 1 step in Figure 6c. This reduction is also evident in the mean distance, which fell from 2 to 1.67 to 1.58 for person A across the three networks.

5.4 Methods

Each aim of this work will further our overall understanding of why epidemic dynamics differ in Cebu City and Mandaue. In Aim 1, we will use RDS data to reconstruct the underlying networks in each city and compare their structures. In Aim 2, identification of transmission clusters will allow us to assess how interconnected infections from Mandaue are to those in Cebu City. Finally, in Aim 3, we bring all these data together to reproduce the epidemic dynamics using an agent-based model.

[5] There are a number of paths to get from A to D. For example A-B-H-K-D or A-F-I-C-D are two possibilities, but there are also several others. (We do not name all possible paths, but only make the point that a minimum of 4 steps is required.)

5.4.1 Aim 1: Compare network structure and behaviors in Cebu City and Mandaue to identify important differences that correlate with differences in HIV prevalence.

To compare network structures in the two cities, we simulate the true underlying networks using an exponential random graph model approach [70] and examine whether the network structures in the two cities are different. Before we set up the simulation, we must first apply RDS weights to our sample, to account for dependence in sampling. We use these weights to estimate degree distribution, homophily and clustering in the network. These estimates are inputs for the exponential random graph (ERG) model, which we use to simulate a network. The network parameters in the ERG model are then updated and adjusted to fit with measures of homophily and clustering on the observed network. A discussion of the strategy and logic behind each of these methods follows here.

RDS-weighted estimates of degree, clustering and homophily

This work uses the RDS survey of IDUs conducted in two cities in Cebu province in 2013. Because each sample is not an independent random sample, we must weight estimates based on their degree. In its most general form, the RDS-corrected estimate of a population mean of some characteristic f would be:

$$\hat{\mu}_f = \sum_{i=1}^{n} \frac{f(X_i)}{d_i} \Big/ \sum_{i=1}^{n} \frac{1}{d_i}$$

where d_i refers to each person's degree, and degree is proportional to one's probability of being sampled. The operational and implementation details of respondent-driven sampling were described above in Section 5.2.3. This section addresses some of the technical and methodological considerations necessary with this method.

Assumptions: Use of the weights presented in the equation above requires several assumptions:

1) Initial sample (seeds) are sampled independently and their selection probability is proportional to their degree;
2) All relationships are symmetric: if X names Y as a tie, Y would also name X as a tie;
3) Participants recruit at random from their contacts: in other words, if someone names 5 friends, then each person among those 5 has an equal probability of being selected for a coupon;
4) No refusal;
5) Sampling with replacement, which is approximately true when sample size is much smaller than the true population size;
6) Number of recruits per participant is non-differential with respect to other covariates; and
7) Respondents accurately report their degree or number of partners.

These assumptions are difficult to confirm, but we assume that violations are not sufficiently large to alter the validity of our estimates.

Variance estimates are also of importance and there is some dispute in RDS literature about the most appropriate choice. Simulations and comparisons with empirical data suggest that variance estimators initially introduced with RDS [66] underestimate true variability, and may fail to capture the true value of some network statistics [71,72]. Improved variance estimators have been proposed and appear to show more realistic results [73], so these estimates will be used when variance calculations are needed.

We calculate RDS-adjusted estimates for three parameters in each of the two cities, which serve as inputs to the network simulations:

- Degree distribution, which depicts the frequency at each level of number of partners. These data are drawn from the questionnaire, which asks the number of injecting partners the respondent has.
- Homophily with respect to several demographic and behavioral characteristics, which include age, gender (m/f), visits a shooting gallery (y/n), source of needles, and drug of choice.

- Transitivity, which will be measured using the geometrically weighted edgewise shared partner (GWESP) index:

$$GWESP = e^{\alpha} \sum_{i=1}^{n-2} \{1 - (1 - e^{-\alpha})^i\} p_i$$

where α sets the rate of decay for each additional partner and p_i is the number of connected dyads who have individual i in common.

This indicator is also described as the proportion of triads (groups of 3) that are closed triangles [21,74]. The data required to calculate this index can be extracted from follow-up questionnaires, which are conducted when participants return to collect their secondary incentive and are asked about their recruits.

Finally, the distribution of ego-network configurations will be used to validate model fit and ensure our model is a simulation of the true underlying network structure. Configurations will be identified using the follow-up questionnaires and then weighted. These will serve as inputs into various steps of the network simulation, using exponential random graph (ERG) and simulation models described in the next sections of this approach.

Exponential random graph model

Because respondent-driven sampling (RDS) methods use the network for recruitment, we exploit the integrated information about localized ties to describe the underlying network structure. These data elucidate a part of the network, but the data are still missing on other parts of the network. Typically, it is difficult to analyze networks when data are not complete (that is, when we have interviewed a small portion of people who are truly part of a network). If a single node is missed, it could significantly affect all network measures (especially measures such as clustering and social distance). Exponential random graph (ERG) models can be used to infer and simulate the "true" network structure from data on a sample of network members [49].

Exponential random graph (ERG) models are a class of models that describe ties (or connections) between people (based on the network structures and attributes of the actors in them). These models employ a statistical method that accounts for the inherent dependent structure of networks. In the past, ERG models were used to describe important factors on fully enumerated networks (i.e., networks with complete data), but these models can also be used on sampled data to estimate network parameters, which may then be used to simulate complete networks [73]. The studies that have applied ERG models in this way often have additional information about higher-ordered network structures and clustering – or they do not have higher-order clustering at all (as is the case of many sexual networks).[6]

Injecting risk networks are highly transitive, so network members (actors or nodes) are more likely to cluster together, which will help to constrain the parameter space for fitting these models. We will describe crossties among part of our sample, but the network information is still incomplete, which dictates the need for a network simulation. From the simulation, we can compare whether network statistics (e.g., longest geodesic path or number of components) differ across the two cities' networks.

ERG models estimate the probability or likelihood of two people in a network having a particular relationship, adjusting for all other existing ties in the network. In its general form:

[6] One important network characteristic that features in social (friendship) networks but not in sexual networks is transitivity. An example of transitivity would be if two female partners who had sex with the same male partner were more likely to be partners. Clustering might be if two female partners had sex with the same male partner, they are more likely to share another male partner, which we also would not expect to be the case.

$$\Pr(\mathbf{Y} = \mathbf{y} | n \text{ actors}) = \left(\frac{1}{\kappa}\right) \exp\left\{\sum_A \eta_A g_A(\boldsymbol{y})\right\}$$

where **Y** is a random network tie that may be present (y=1) or absent (y=0), and **y** is the network that we observe. The function $g_A(\boldsymbol{y})$ refers to possible network structures (such homophily or transitivity) that may explain the likelihood of a tie; and η_A is a coefficient related to the association of each network measure with the likelihood of a tie. The $1/\kappa$ is a normalizing constant. Goodreau and colleagues (2009) further show that this model could also be described as:

$$\text{logit}(\mathrm{P}(\mathrm{Y_{ij}} = 1 \big| \, n \text{ actors}, \mathrm{Y_{ij}^c})) = \sum_A \eta_A \delta g_A(y)$$

where $\mathrm{Y_{ij}^c}$ represents all other ties in the network, and $\delta \mathrm{g_A}(\boldsymbol{y})$ is a correction factor. This description makes more explicit the idea that ERGMs are able to test whether ties occur by chance, even after adjusting for the tendency for ties to occur in the presence of other underlying network structures. In other words, ERGMs adjust for the dependency structure of the networks, which is implied when particular features are present.

Assortative mixing and transitivity were previously described in Section 5.3, but let us further explore the dependence of different features of the network. Consider a network with high homophily. This implies that people with similar characteristics or interests tend to be connected. For example, Alan injects "speedball" (a mix of heroin and methamphetamines) with two of his friends – Daniel and Geoff. If Daniel and Geoff also inject together, this suggests high transitivity in the network. However, if the number of speedball injectors is very small and assortative mixing is high, then we would expect that Daniel and Geoff are also friends who inject together simply by chance. Therefore, if we specify transitivity as a separate parameter in our ERG model, its coefficient may be very small, even when it appears that the network is highly transitive. It is important to consider how these effects interact before we build the ERG model.

Running the Simulation

Several steps are involved in simulating these networks, as described by Smith [70]. This procedure could be considered in three phases. First, we describe each network's characteristics and simulate a random network from the observed data. Then, we estimate an ERG model and update the network models to reproduce the measures of homophily observed in the network. Finally, this updated simulation is calibrated to match the clustering, as measured by the GWESP index in the data. Each step is outlined below:

(A) Setting initial model parameters

The first step is to set the main parameters for the simulation. To do this, we set the size of the IDU population (N) in each city based on prior population estimates. We then calculate the degree distribution, and higher-order network configurations from our sampled data. Next, we create a random network of size N. People in this hypothetical network are assigned demographic and behavioral characteristics that correspond to people with comparable degree in the observed data. Therefore, we should have the correct degree distribution and distribution of covariates in our simulated population.

(B) Fitting models and the initial network

We then specify an ERG model, where the probability of a network tie occurring is based on the demographic and behavioral variables that determine homophily, and a "node covariate" variable, which

is the degree in the initial random network constructed in part (A).[7] Homophily will be defined by those measured characteristics important to formation of a tie between two injectors, which may include factors such as age, location, drug of choice, and injecting practices (such as visiting shooting galleries). We then set initial coefficients and constraints on the model, which will reduce the parameter space, and improve efficiency of model estimation. The results of this model are used to adjust our simulated network to improve its fit to homophily statistics.

Starting from the random network structure, we simulate a new model that incorporates the ERG model coefficients (for assortative mixing effects) estimated above. In the next step, this new model is adjusted to match homophily and clustering characteristics from our observed data.

(C) Simulating the networks to fit the data
Although homophily matches the ERG coefficient estimates, it may not exactly match our data. Therefore, we will further update the homophily estimates by comparing results from our simulated network to the RDS-weighted homophily estimates. If assortative mixing in our simulation is too weak, the model coefficients in the simulation should be increased. Similarly, if mixing effects imply too much homophily, then model coefficients should be decreased to better fit the model. Since our initial estimates are based on draws from the observed data, we expect minimal adjustment will be required in this step.

Once we have finished tuning the homophily estimates, we will adjust the coefficient on the clustering parameter (GWESP), which can take on different values depending on the distribution of different local network configurations. After updating this aspect of the model, we will use a chi-square test to estimate how well the model fits the observed (weighted) distribution of configurations. A good fit suggests that our model coefficients and specification predict a structure that could give rise to the observed data.

This process will be implemented to simulate separate networks in Cebu City and Mandaue. Once the ERG models are completed, we will compare network structures by considering model coefficients on our measures for clustering and different types of mixing. A high coefficient on the clustering parameter suggests that people are well connected and infection may spread more efficiently through its network members. Therefore, we would expect a higher clustering coefficient in Cebu City compared to Mandaue, which would provide an explanation for how infection spread so rapidly in this city. These differences will be further tested in the transmission model (Aim 3).

Limitations
This work is limited in its dependence on assumptions about RDS data and on correct ERG model specification. The highly interdependent structure of networks creates high-order constraints that allow us to model the true underlying network even in the absence of complete data. It is this interdependence that also provides additional structure that can improve the robustness of estimation and validation.

The largest limitation of this study is the cross-sectional nature of the data, which precludes any causal inference we might make about the relationship between injecting ties, related HIV risk behaviors, and the transmission of blood-borne disease. The results of this analysis will serve as inputs to the network and HIV transmission models (Aim 3), and if we can replicate the true transmission dynamics observed, we should feel confident that our network structures are close to truth. Moreover, we can use these transmission models to run scenarios that change network structures or behaviors, holding the other factor constant, to describe their relative effects on transmission dynamics.

5.4.2 Aim 2: Examine phylogenetic clustering of HIV infections detected in Cebu and Mandaue.

Analysis of viral sequences will be used to examine whether transmission of infection occurred between the two cities or primarily within each city. We will analyze HIV nucleotide sequences from specimens collected during HIV surveillance of IDUs to assess how related the infections are across these two cities.

[7] This variable is included because we know that this initial model has a degree distribution and characteristics that match the weighted data; therefore, we want to ensure that our ERG model does not stray too far from this distribution.

We hypothesize that infections within a city are more closely related to each other than to those from the other city. Since these data are tracked with unique IDs, we will be able to link the viral population within each patient to his or her demographic and behavioral characteristics.

We will obtain next generation sequencing (NGS) data from the Hawaii Center for AIDS (HCFA), which is conducting the sequencing on behalf of the Philippine Department of Health. These data provide greater detail on the viral diversity within each person. The mutation rate for HIV ranges from 1 per 100,000 to 1 per 2000 base pair generation [69,75]. These mutations either make the virus stronger or weaker. A weak strain has poor viral fitness; it will not be replicated and will eventually die off. Strains with high viral fitness usually have a selective advantage in terms of drug resistance, which is why they may be more prevalent in a person, depending on the virus that was transmitted, duration of infection and treatment history. Therefore, the use of NGS data could be helpful in identifying small differences between people with otherwise similar HIV genetic profiles.

The gain in read depth in NGS comes at the cost of shorter read length. While earlier studies of HIV transmission clusters read long segments of the *pol* region and are often supplemented by sequences of *gag* and *env* regions, these data are targeted to obtain in-depth nucleotide sequences for 438 base pairs in the p51 RT region of the *pol* gene (Figure 8).

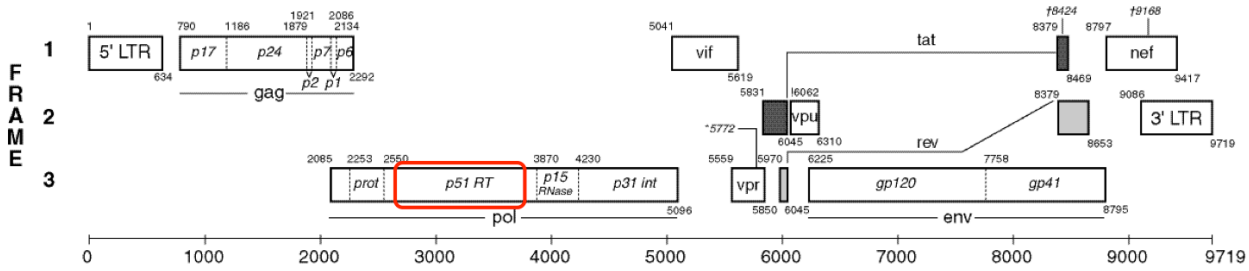


Figure 8. Map of HIV-1 and the p51 RT region of *pol* gene that is the target for sequencing (from: http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html)

Data Cleaning

Next-generation sequencing provides up to 100,000 observations (or "reads") per sample. Each read is a sequence of about 500 base pairs: the first 50 encode the patient ID, and the remaining 450 in the *pol* region of the HIV-1 genome (targeting p51 region, HXB2 2667-3593). Once sequencing data are collected, several additional steps are taken to ensure quality control and prepare the data for analysis. Very short chains (<100 base pairs) are removed from the sample, and gaps are removed from the longer chains, and poor quality sequences (as measured by phred[8] scores) are removed. The reverse complement of 3' sequences is computed to get 5' reads, so that alignment is of the same strand.

After completing these quality control measures, each sequence is pushed into a fasta alignment, grouped by patient and amplicon. We expect 1000-5000 sequences per patient for alignment and analysis. We first align the sequences within each patient-sample by comparing them to a referent multiple sequence alignment using BLAT software [78]. The software output will report details on mutations, including the base and its location, by patient-amplicon. These data will not be used for trees, but are maintained for considering in-depth analysis within transmission clusters.

We will use these data to identify transmission clusters by building phylogenetic trees defined by differences in the DNA sequences across patients. To build the trees, we will extract a single consensus sequence per patient, based on the dominant base present at each position. We will use the same BLAT

[8] Phred scores are a general measure of sequencing quality that is conducted within the sequencing instrument's software. See [76,77] for details on the QC algorithm.

alignment, and output all the sequences by patient and amplicon. Then we will use the consensus-generating tool from Los Alamos laboratories[9] to extract a single sequence per patient. These consensus sequences will be combined with reference sequences from the same surveillance round among MSM, other sequences from the Philippines, and from a BLAST search of related HIV sequences available in GenBank. We will use MEGA software [79] to generate alignments across patients and build trees. Subtypes and intra-subtype recombination will be identified using Subtype Classification Using Evolutionary Algorithms[10]. Mutations associated with major drug resistance will also be evaluated with the Stanford HIV drug resistance database.

We will apply both Neighbor Joining (NJ) and Maximum Likelihood (ML) methods to build trees. Analyses for each of these methods are discussed below. The use of different methodological approaches to identify transmission clusters should support the reliability of our results.

Neighbor Joining (NJ)

The NJ method constructs phylogenetic trees based on pairwise genetic distance, which is measured in terms of the number of bases that differ between all pairs of sequences. A correction is applied for multiple substitutions so that the genetic distance reflects the expected mean number of changes per site [80]. The phylogenetic tree constructed is the relationship between these distance values. NJ methods are fast and proven efficient in finding accurate tree topologies [61]. To estimate the statistical robustness of our results, we will also calculate bootstrap estimates of our results. Only those branches that occur in a very high proportion of replicates will be considered valid in our analysis.

Maximum Likelihood (ML) method

We will also use a Maximum Likelihood (ML) analysis to test our hypothesis and assess whether results are consistent over different methods. Based on the generalized time-reversible evolutionary model (discussed below), we derive a likelihood function and estimate parameters values that give the model that maximizes the probability of the data. If necessary, we will seed the fitting parameters based on those used in previous studies in Cebu [68,81] or based on results from the neighbor joining method, to reduce the parameter space and ensure model convergence.

Evolutionary Model

For either NJ or ML method, we must assume an evolutionary model that will determine the genetic distance between two sequences where all mutations occur at the same site or set of sites. For this work, we will apply a general time-reversible (GTR) model of evolution, which is commonly used for its flexibility: this model places only two constraints on the 10 fitting parameters. This means that most parameters can be fit based on available data. Specifically, the GTF model allows for variation in eight parameters: five estimates of nucleotide substitution rates and three estimates of base prevalence in the sequence,[11] as described in matrix Q:

$$\boldsymbol{Q} = \begin{pmatrix} a\pi_C - b\pi_G - c\pi_T & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & a\pi_A - d\pi_G - e\pi_T & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & b\pi_A - d\pi_C - f\pi_T & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & c\pi_A - e\pi_C - f\pi_G \end{pmatrix}$$

where $\pi_A, \pi_C, \pi_G, \pi_T$ are the proportion of base frequencies such that they sum to 1, and their coefficients (a-f) refer to the based substitution probabilities, which also sum to 1.

[9] Available at: http://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html
[10] Available at: http://www.datamonkey.org/dataupload_scueal.php
[11] Note that here we say only five of the substitution rates are free parameters because once five are set, the sixth is implied because of the constraints: ($a + b + c + d + e + f = 1$); similarly, with prevalence parameters, estimating three determines the fourth ($\pi_A + \pi_C + \pi_G + \pi_T = 1$).

This model assumes that these substitution probabilities are equal (that is, the rate of a substitution of A for G is equal to the substitution of G for A). This assumption means that we cannot distinguish between forwards and backwards substitution (time reversible), which precludes any attempt to make causal inference about the direction of transmission.

Cluster Identification

There is no good consensus on how to define a phylogenetic cluster. Studies provide a range of thresholds, whether based on genetic distance (0.015 – 0.030 mean distance) or clade support, measured by 70-98% bootstrap or 90-100% Bayesian posterior probabilities [82]. Since HIV infection in our study population appears relatively homogeneous, we will choose a high threshold (i.e., short distance) of ≤0.015 mean genetic distance with 95% bootstrap. According to the genetic distance criteria, most IDU infections detected in the previous surveillance rounds would be part of a single large cluster [68], as we would expect.

Limitations

We may be limited by the success rate of viral amplification, which may limit sample size. Blood and plasma are collected in the Philippines and then separated for HIV, HCV and Syphilis testing before being shipped to the sequencing laboratory at the Hawaii Center for HIV/AIDS (HCFA). In some samples, most of the virus may have degraded, precluding any possibility of amplification or sequencing. Although the amplification success rate of 22% is quite low, there is no reason to believe that degradation is differential with respect to epidemic, and we will use these data to broadly describe patterns in transmission across the two cities. Therefore, we will continue the analysis with caution and consider our conclusions in light of these limitations.

Since the data are collected primarily for public health (and not research) purposes, the sequencing data were targeted to look in-depth along those regions of the genome that carry clinically important mutations. As a result, if the consensus sequences are short, there may not be sufficient phylogenetic signal to detect differences between two or more different patient sequences. If such a situation arises, we may analyze the NGS reports on common patterns in minority mutations to distinguish between them. Hospital records (as reported at the central level) and expert opinion indicate that no IDUs had initiated ART at the time of the 2013 surveillance, so we do not expect convergent evolution to be a significant concern in this population.

Another limitation is that our data are missing information about duration of infection. Most of the IDUs who participated were unaware of their HIV status, so it may be difficult to distinguish between genetic distance resulting from long-term evolution and infections that have truly distinct sources of infection. One way to estimate duration of infection is to use results from BED assays, which are used to identify infections that occurred within the last one year. This diagnostic test was conducted in the Philippines as part of the surveillance testing protocol.

5.4.3 Aim 3: Construct a mathematical transmission model to compare the relative contributions of the network and individual injecting behaviors to the epidemic dynamics in Cebu and Mandaue.

To test our hypothesis that different network structures explain the different observed HIV prevalence, we will simulate HIV transmission across individuals in each network. We will compare the equilibrium prevalence and time to reach equilibrium by network to look for differences by city. This simulation applies an agent-based model, which will assign behaviors and network descriptions to each individual node (or actor) along each network. The basic parameters of this model are based on a modified design of the agent-based models of heterosexual networks [49], which we describe here.

Basic Model Equation

This model follows a basic susceptible-infected (SI) structure, since people who are HIV-infected do not recover and are not vulnerable to infection (since we do not model super-infection). We model

transmission, such that for any sero-discordant dyad (pair) of injecting drug users i and j, the probability of transmission per exposure event (β_x) can be expressed as:

$$\beta_x = 1 - (1 - \beta_{ij})$$

where β_{ij} is the probability of transmission per shared injection. Then, the cumulative probability of transmission for a specific pair of injecting drug users (β_p) over the duration T (in weeks) across this pair (p) is:

$$\beta_p = 1 - \left(1 - \beta_{ij}\right)^{nT}$$

where n is the number of shared injections in the dyad per week and β_{ij} is as defined above. Here, the parameter n reflects two aspects of risk behavior: first, the frequency of injection, which may be influenced by choice and accessibility of drugs and needles; and second, the prevalence of injection sharing, which reflects the availability of clean needles as well as social influences (e.g., trust, age or power differential, etc.). At each time step of the model, a random draw is taken from this probability distribution to simulate the spread of infection across the network.

The simulation is run across dyads, but conceptually, we can also describe how these transmission probabilities may translate to an individual: specifically, an HIV-uninfected IDU who shares needles with k HIV+ partners will have a risk of infection (β_i) over one iteration period of T weeks that can be calculated as:

$$\beta_i = 1 - \prod_{j=1}^{k} \left[\left(1 - \beta_{ij}\right)^{n_j T}\right]$$

In this equation, all parameters are the same as described above. Here, we take 1 minus the product of the probability of no infection across all sero-discordant partnerships to estimate the cumulative risk (across many partners) in a single iteration of the model.

Now, we describe some of the parameters in greater detail.

Parameter Definition and Data Sources

Three parameters define the injecting risk associated with any HIV-discordant IDU dyad (as described above): (i) the biological transmission probability (β_{ij}); (ii) the frequency of injection (n); and (iii) the duration of the partnership (T). In this section, we provide more detail on data sources and assignment procedure for parameter values.

Biological probability of transmission (β_{ij}): It is difficult to accurately estimate the risk of HIV transmission per needle stick exposure. Transmission will depend upon the type of needle and the nature of sharing. For example, low dead-space syringes retain less residual blood or drug, and therefore have a lower probability of transmission [83]. Similarly, sharing drugs or cookers have significantly lower rates of transmission than sharing needles and syringes. It is estimated that HIV transmission will occur somewhere between 2 to 6 per 1,000 needle stick exposures [13]. We will adjust this parameter to fit results to our observed data.

Weekly number of shared injections (n): How frequently an IDU shares injections with someone determines his potential exposure-level with that particular partner. This parameter is measured by creating a joint distribution of frequency of injecting and prevalence of sharing, both of which are asked in behavioral surveillance. While these data are collected at the individual level, the model parameter is expressed in reference to a specific dyad. To assign individual-level observed data to a pair of individuals, we first construct the joint distribution of the number of injections and prevalence of sharing per week, based on reported surveillance. Then, the model simulation will take a new random draw from this distribution for each discordant pair at each time point.

Partnership duration (T): Since we have a single estimate of the network, we will assume the duration of each injecting pair will be maintained for the full simulation. We may perform sensitivity analyses by adjusting this parameter to see its effect on how quickly the epidemic spreads across each city.

Simulating the epidemic

Once each network configuration is established, we randomly introduce a small number of HIV infections into each network. The number and timing of the introduction of new infections will be informed by the phylogenetic clusters identified in Aim 2. If several distinctly different clusters are identified, then it may be appropriate to introduce a comparable number of new infections to simulate transmission from IDUs in the other city's network. The model will update and repeat the infection pattern every week over the 10-year period (2009-2019) to predict the networks' impacts on the growth of HIV in these two populations. We repeat this simulation 1,000 times in each city, and compare distributions of the results to look for differences across the two cities.

Based on these results, we will assess the validity of our network and transmission parameters by attempting to replicate the epidemic dynamics observed through surveillance in 2009-2013. The epidemic in Cebu City showed a dramatic increase in less than a year while data from Mandaue suggested slower growth. If our models do not replicate these transmission dynamics, we may question our description of networks or behaviors. On the other hand, if we can successfully and confidently reconstruct the epidemic dynamics in these two cities, we may be able to use the models to run scenarios and estimate the relative contributions of injecting behaviors and network structures on the epidemics.

Limitations

The major limitation of this model is its dependence on the validity of the networks simulated in Aim 1. If these networks are incorrect, our model may estimate a very different epidemic scenario than the one we observe. We expect that these network models will be robust, as they are calibrated to several measures to ensure they reproduce the observed data and accurately represent the population. However, if the epidemic dynamics are quite different between the two cities and if they cannot be replicated in our models, this may be a strong indication that our network model is incomplete.

The use of a static network is probably not realistic for modeling transmission dynamics for durations longer than one year. Injecting drug use practices are not life-long. People may choose injecting partners according to who is available or has money to purchase drugs at a given opportunity. Even within smaller networks of high trust, people will enter and exit the networks as their circumstances change. We recognize that the ability to simulate the spread of infection on the network as the network structure itself is changing (with people forming new partnerships and dissolving old ones) is quite important; however, at this point, the data for such simulation are still unavailable.

5.4.4 Statistical Power and Sample Size

Because we are modeling networks from each city separately, we will not conduct formal statistical tests to compare them. Instead, we will compare coefficients of our two ERG models to assess whether mixing or clustering patterns differ between these two cities. The limiting sample size for this project, then, is the RDS data, which we use to construct degree, demographic and behavioral information that will be used to construct the networks.

Power and sample size calculations for RDS data follows the general formula for sample size calculations, with the addition of the design effect parameter. Thus, the calculation for sample size needed to detect a difference in sharing rates in the two cities can be calculated as:

$$n = D\frac{\left[Z_{1-\alpha}\sqrt{\bar{p}(1-\bar{p})} - Z_{\beta}\sqrt{0.5 * p_1(1-p_1) + 0.5 * p_2(1-p_2)}\right]^2}{0.25 * (p_1 - p_2)^2}$$

With the exception of the D parameter, the equation is the standard sample size calculation for a one-sided test to detect a difference in proportions $(p_1 - p_2)$. This D parameter, also called the design effect, should be at least two, but newer studies suggest the need for a design effect of 4 or higher [84,85]. Using a design effect of 2 and assuming 70% sharing in Mandaue, our sample sizes of 457 in Cebu City and 310 in Mandaue have about 93% power to detect a 15% difference in injection sharing in the two cities. For the larger design effect of 4, power drops to 66%.[12] Small differences in proportions will be difficult to detect, but it is difficult to state exactly how this power calculation translates to the reliability of our network simulation. Instead, we will build this uncertainty into the simulations, and conduct sensitivity analyses to see how different shaped distributions may affect our network and transmission model results.

As discussed previously, the phylogenetic analysis in Aim 2 may be limited by the loss of data due to RNA degradation during transport from Cebu to Manila to Hawaii. Currently, 22% of attempted infections were successfully amplified. This success rate is not uncommon in HIV amplification, and we do not expect the missing data to bias our results, though it may reduce our statistical power. Therefore, we should remain cautious and take care not to make too strong conclusions from this analysis.

The transmission model (Aim 3) will simulate the underlying mechanisms for transmission and using this method, we can conclude whether it is the network structures that are important to transmission dynamics. Power is not estimated for this aim, as the test of its validity will be whether the proposed network structures and parameters can replicate the HIV dynamics observed in this community.

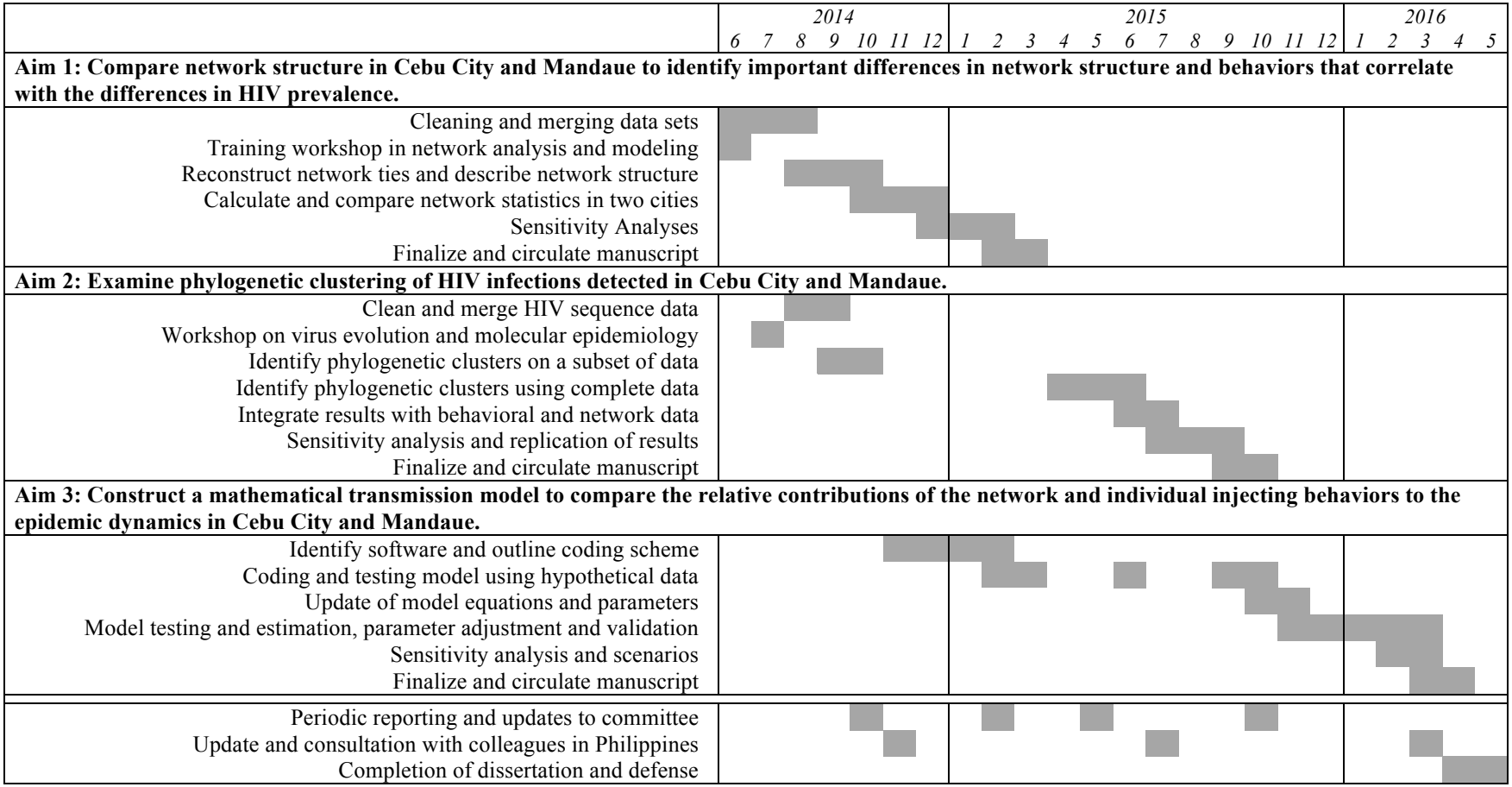
[12] All power calculations were conducted using Stata 13, 'power' command; the design effect was incorporated by reducing the estimated sample size by a factor of the design effect. A design effect of 2, then, was incorporated by halving the sample sizes.

6 Summary of Public Health Impact

The proposed scope of this work, which capitalizes on a combination of network, behavioral and phylogenetic data, is uniquely positioned to explore the influence of network structure on HIV transmission among IDUs. We can use information about network structures to identify people living with HIV and construct or target prevention strategies in two ways. First, if we can reconstruct networks and describe the circumstances that caused their infection, we may use this information to direct and design prevention interventions for similar networks. Second, if we can describe the location of HIV infections in terms of the network positions, we can simulate their impacts on the remainder of the network to identify who is most important to propagating infection and who is most at risk of acquiring it. As data collection continues in this population in the Philippines, we may also be able to test the validity of our tools and results by analyzing subsequent rounds to look for consistencies or changes over time.

7 Timeline

I expect this work to take approximately 2 years, as described below:



8 References

1. UNODC. World Drug Report. 2013.

2. Sarkar S, Das N, Panda S, Naik TN, Sarkar K, Singh BC, et al. Rapid spread of HIV among injecting drug users in north-eastern states of India. Bull Narc. 1993;45(1):91-105.[cited 2013, Sep 29]

3. The Economic Impact of Illicit Drug Use on American Society. The Economic Impact of Illicit Drug Use on American Society. 2011, Apr;

4. 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. Lancet. 2008, Nov 15;372(9651):1733-45.

5. Dutta A, Wirtz A, Stanciole A, Oelrichs R, Semini I, and Cleghorn F. The Global HIV Epidemics among People Who Inject Drugs. World Bank Publications; 2012.

6. Morineau G, Bollen LJ, Syafitri RI, Nurjannah N, Mustikawati DE, and Magnani R. HIV prevalence and risk behaviours among injecting drug users in six indonesian cities implications for future HIV prevention programs. Harm Reduct J. 2012;9(1):37.

7. Des Jarlais DC, Feelemyer JP, Modi SN, Arasteh K, Mathers BM, Degenhardt L, and Hagan H. Transitions from injection-drug-use-concentrated to self-sustaining heterosexual HIV epidemics: patterns in the international data. PLoS One. 2012;7(3):e31227.

8. Saidel TJ, Des Jarlais D, Peerapatanapokin W, Dorabjee J, Singh S, and Brown T. Potential impact of HIV among IDUs on heterosexual transmission in Asian settings: scenarios from the Asian Epidemic Model. International Journal of Drug Policy. 2003;14(1):63-74.

9. Volz E, Frost SDW, Rothenberg R, and Meyers LA. Epidemiological bridging by injection drug use drives an early HIV epidemic. Epidemics. 2010, Sep;2(3):155-64.

10. Vanichseni S, Choopanya K, Des Jarlais DC, Sakuntanaga P, Kityaporn D, Sujarita S, et al. HIV among injecting drug users in Bangkok: the first decade. International Journal of Drug Policy. 2002;13(1):39-44.

11. Redefining AIDS in Asia: Crafting an Effective Response. Redefining AIDS in Asia: Crafting an Effective Response. 2008, Mar 26;

12. Wiebel WW, Jimenez A, Johnson W, Ouellet L, Jovanovic B, Lampinen T, et al. Risk behavior and HIV seroincidence among out-of-treatment injection drug users: a four-year prospective study. J Acquir Immune Defic Syndr Hum Retrovirol. 1996, Jul;12(3):282-9.

13. Baggaley RF, Boily M-C, White RG, and Alary M. Risk of HIV-1 transmission for parenteral exposure and blood transfusion: a systematic review and meta-analysis. AIDS. 2006, Apr 4;20(6):805-12.

14. Boily M-C, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, and Alary M. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. Lancet Infect Dis. 2009, Feb;9(2):118-29.

15. Friedman SR, Jose B, Deren S, Des Jarlais DC, and Neaigus A. Risk factors for human immunodeficiency virus seroconversion among out-of-treatment drug injectors in high and low seroprevalence cities. The National AIDS Research Consortium. Am J Epidemiol. 1995, Oct 15;142(8):864-74.

16. Nicolosi A, Leite ML, Musicco M, Molinari S, and Lazzarin A. Parenteral and sexual transmission of human immunodeficiency virus in intravenous drug users: a study of seroconversion. The Northern Italian Seronegative Drug Addicts (NISDA) Study. Am J Epidemiol. 1992, Feb 1;135(3):225-33.

17. Patrick DM, Strathdee SA, Archibald CP, Ofner M, Craib KJ, Cornelisse PG, et al. Determinants of HIV seroconversion in injection drug users during a period of rising prevalence in Vancouver. Int J STD AIDS. 1997, Jul;8(7):437-45.

18. Marmor M, Des Jarlais DC, Cohen H, Friedman SR, Beatrice ST, Dubin N, et al. Risk factors for infection with human immunodeficiency virus among intravenous drug abusers in New York City. AIDS. 1987, May;1(1):39-44.

19. Grund JP, Friedman SR, Stern LS, Jose B, Neaigus A, Curtis R, and Des Jarlais DC. Syringe-mediated drug sharing among injecting drug users: patterns, social context and implications for transmission of blood-borne pathogens. Soc Sci Med. 1996, Mar;42(5):691-703.

20. Deren S, Beardsley M, Coyle S, Singer M, and Kang SY. HIV risk behaviors among injection drug users in low, medium, and high seroprevalence communities. AIDS and Behavior. 2001;5(1):45-50.

21. Valente TW. Social Networks and Health: Models, Methods, and Applications. New York: Oxford University Press; 2010. [cited 2013, Dec 3]

22. Fichtenberg CM, Muth SQ, Brown B, Padian NS, Glass TA, and Ellen JM. Sexual network position and risk of sexually transmitted infections. Sex Transm Infect. 2009, Dec;85(7):493-8.

23. Friedman SR, Neaigus A, Jose B, Curtis R, Goldstein M, Ildefonso G, et al. Sociometric risk networks and risk for HIV infection. Am J Public Health. 1997;87(8):1289-1296.

24. Woodhouse DE, Rothenberg RB, Potterat JJ, Darrow WW, Muth SQ, Klovdahl AS, et al. Mapping a social network of heterosexuals at high risk for HIV infection. AIDS. 1994, Sep;8(9):1331-6.

25. Liljeros F, Edling CR, Amaral LA, Stanley HE, and Aberg Y. The web of human sexual contacts. Nature. 2001, Jun 21;411(6840):907-8.

26. Watts DJ, and Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998, Jun 4;393(6684):440-2.

27. Amaral LA, Scala A, Barthelemy M, and Stanley HE. Classes of small-world networks. Proc Natl Acad Sci U S A. 2000, Oct 10;97(21):11149-52.

28. Rothenberg RB. Commentary: sampling in social networks. Connections. 1995;18(1):104-110.

29. Valente TW, Coronges K, Lakon C, and Costenbader E. How correlated are network centrality measures? Connections (Toronto, Ont.). 2008;28(1):16.

30. Wu Z, Detels R, Zhang J, Duan S, Cheng H, Li Z, et al. Risk factors for intravenous drug use and sharing equipment among young male drug users in Longchuan County, south-west China. AIDS. 1996, Aug;10(9):1017-24.

31. Aitken CK, Higgs P, and Bowden S. Differences in the social networks of ethnic Vietnamese and non-Vietnamese injecting drug users and their implications for blood-borne virus transmission. Epidemiol Infect. 2008, Mar;136(3):410-6.

32. Fuller CM, Vlahov D, Latkin CA, Ompad DC, Celentano DD, and Strathdee SA. Social circumstances of initiation of injection drug use and early shooting gallery attendance: implications for HIV intervention among adolescent and young adult injection drug users. J Acquir Immune Defic Syndr. 2003, Jan 1;32(1):86-93.

33. Neaigus A, Friedman SR, Curtis R, Des Jarlais DC, Furst RT, Jose B, et al. The relevance of drug injectors' social and risk networks for understanding and preventing HIV infection. Soc Sci Med. 1994, Jan;38(1):67-78.

34. Subbarao S, Vanichseni S, Hu DJ, Kitayaporn D, Choopanya K, Raktham S, et al. Genetic characterization of incident HIV type 1 subtype E and B strains from a prospective cohort of injecting drug users in Bangkok, Thailand. AIDS Res Hum Retroviruses. 2000, May 20;16(8):699-707.

35. Beyrer C, Razak MH, Lisam K, Chen J, Lui W, and Yu XF. Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia. AIDS. 2000, Jan 7;14(1):75-83.

36. Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, Serwadda D, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. AIDS. 2009, Sep 10;23(14):F9-F17.

37. Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, and Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A. 2007, Nov 20;104(47):18566-70.

38. Bello G, Aulicino PC, Ruchansky D, Guimarães ML, Lopez-Galindez C, Casado C, et al. Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. Retrovirology. 2010;722.

39. Tee KK, Pybus OG, Li X-J, Han X, Shang H, Kamarulzaman A, and Takebe Y. Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08_BC and 07_BC in Asia. J Virol. 2008, Sep;82(18):9206-15.

40. Liao H, Tee KK, Hase S, Uenishi R, Li X-J, Kusagawa S, et al. Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. Virology. 2009, Aug 15;391(1):51-6.

41. Lewis F, Hughes GJ, Rambaut A, Pozniak A, and Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med. 2008, Mar 18;5(3):e50.

42. Pilcher CD, Wong JK, and Pillai SK. Inferring HIV transmission dynamics from phylogenetic sequence relationships. PLoS Med. 2008, Mar 18;5(3):e69.

43. Oster AM, Pieniazek D, Zhang X, Switzer WM, Ziebell RA, Mena LA, et al. Demographic but not geographic insularity in HIV transmission among young black MSM. AIDS. 2011, Nov 13;25(17):2157-65.

44. Hurt CB, and Dennis AM. Putting It All Together: Lessons From the Jackson HIV Outbreak Investigation. Sex Transm Dis. 2013, Mar;40(3):213-5.

45. Revenga A. The economics of effective AIDS treatment : evaluating policy options for Thailand. Washington, DC: World Bank; 2006.

46. Foss AM, Watts CH, Vickerman P, Azim T, Guinness L, Ahmed M, et al. Could the CARE-SHAKTI intervention for injecting drug users be maintaining the low HIV prevalence in Dhaka, Bangladesh? Addiction. 2007, Jan;102(1):114-25.

47. Powers KA, Ghani AC, Miller WC, Hoffman IF, Pettifor AE, Kamanga G, et al. The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study. Lancet. 2011, Jul 16;378(9787):256-68.

48. Granich RM, Gilks CF, Dye C, De Cock KM, and Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet. 2009, Jan 3;373(9657):48-57.

49. Morris M, Kurth AE, Hamilton DT, Moody J, and Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. American Journal of Public Health. 2009;99(6):1023.

50. Marshall BDL, Paczkowski MM, Seemann L, Tempalski B, Pouget ER, Galea S, and Friedman SR. A complex systems approach to evaluate HIV prevention in metropolitan areas: preliminary implications for combination intervention strategies. PLoS One. 2012;7(9):e44833.

51. Rolls DA, Daraganova G, Sacks-Davis R, Hellard M, Jenkinson R, McBryde E, et al. Modelling hepatitis C transmission over a social network of injecting drug users. J Theor Biol. 2012, Mar 21;29773-87.

52. Latkin C, Mandell W, Vlahov D, Knowlton A, Oziemkowska M, and Celentano D. Personal network characteristics as antecedents to needle-sharing and shooting gallery attendance. Social Networks. 1995;17(3):219-228.

53. Friedman SR, Neaigus A, Jose B, Curtis R, and Des Jarlais D. Networks and HIV risk: An introduction to social network analysis for harm reductionists. International Journal of Drug Policy. 1998;9(6):461-469.

54. Prybylski D, Acharya LB, Tuladhar SM, Dhungel N, Gautam BR, McPherson J, et al. Using Surveillance Data to Evaluate a Large-Scale HIV Highway Intervention Targeting Female Sex Workers in the Terai Region of Nepal. jHASE-Journal of HIV/AIDS Surveillance & Epidemiology. 2011;3(1):

55. Malekinejad M, Johnston LG, Kendall C, Kerr LRFS, Rifkin MR, and Rutherford GW. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. AIDS Behav. 2008, Jul;12(4 Suppl):S105-30.

56. Robins G, and Morris M. Advances in exponential random graph (p*) models. Social Networks. 2007;29(2):169-172.

57. Robins G, Pattison P, Kalish Y, and Lusher D. An introduction to exponential random graph (p*) models for social networks. Social Networks. 2007, May;29(2):173-191.

58. Nguyen L, Hu DJ, Choopanya K, Vanichseni S, Kitayaporn D, van Griensven F, et al. Genetic analysis of incident HIV-1 strains among injection drug users in Bangkok: evidence for multiple transmission clusters during a period of high incidence. J Acquir Immune Defic Syndr. 2002, Jun 1;30(2):248-56.

59. de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, et al. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature. 2006, Dec 14;444(7121):836-7.

60. Dennis AM, Hué S, Hurt CB, Napravnik S, Sebastian J, Pillay D, and Eron JJ. Phylogenetic insights into regional HIV transmission. AIDS. 2012, Sep 10;26(14):1813-22.

61. Leitner T, Escanilla D, Franzén C, Uhlén M, and Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci U S A. 1996, Oct 1;93(20):10864-9.

62. Dennis AM, Murillo W, de Maria Hernandez F, Guardado ME, Nieto AI, Lorenzana de Rivera I, et al. Social network-based recruitment successfully reveals HIV-1 transmission networks among high-risk individuals in El Salvador. J Acquir Immune Defic Syndr. 2013, May 1;63(1):135-41.

63. De P, Cox J, Boivin J-F, Platt RW, and Jolly AM. The importance of social networks in their association to drug equipment sharing among injection drug users: a review. Addiction. 2007, Nov;102(11):1730-9.

64. Friedman SR, de Jong W, Rossi D, Touzé G, Rockwell R, Des Jarlais DC, and Elovich R. Harm reduction theory: users' culture, micro-social indigenous harm reduction, and the self-organization and outside-organizing of users' groups. Int J Drug Policy. 2007, Mar;18(2):107-17.

65. Rhodes T, Singer M, Bourgois P, Friedman SR, and Strathdee SA. The social structural production of HIV risk among injecting drug users. Soc Sci Med. 2005, Sep;61(5):1026-44.

66. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Social problems. 1997;174-199.

67. Paraskevis D, Nikolopoulos G, Fotiou A, Tsiara C, Paraskeva D, Sypsa V, et al. Economic Recession and Emergence of an HIV-1 Outbreak among Drug Injectors in Athens Metropolitan Area: A Longitudinal Study. PLoS One. 2013;8(11):e78941.

68. Telan EFO, Samonte GMJ, Palaypayon N, Abellanosa-Tac-An IP, Leaño PSA, Tsuneki A, and Kageyama S. Possible HIV transmission modes among at-risk groups at an early epidemic stage in the Philippines. J Med Virol. 2013, Dec;85(12):2057-64.

69. Castro-Nallar E, Pérez-Losada M, Burton GF, and Crandall KA. The evolution of HIV: inferences using phylogenetics. Mol Phylogenet Evol. 2012, Feb;62(2):777-92.

70. Smith JA. Macrostructure from Microstructure: Generating Whole Systems from Ego Networks. Sociological Methodology. 2012, Aug 1;42(1):155-205.

71. Goel S, and Salganik MJ. Respondent-driven sampling as Markov chain Monte Carlo. Stat Med. 2009, Jul 30;28(17):2202-29.

72. McCreesh N, Frost SDW, Seeley J, Katongole J, Tarsh MN, Ndunguse R, et al. Evaluation of respondent-driven sampling. Epidemiology. 2012, Jan;23(1):138-47.

73. Gile KJ, and Handcock MS. Respondent-Driven Sampling: An Assessment of Current Methodology. Sociological methodology. 2010;40(1):285.

74. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, et al. Networks and the epidemiology of infectious disease. Interdiscip Perspect Infect Dis. 2011;2011284909.

75. Abram ME, Ferris AL, Shao W, Alvord WG, and Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J Virol. 2010, Oct;84(19):9864-78.

76. Ewing B, Hillier L, Wendl MC, and Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998, Mar;8(3):175-85.

77. Ewing B, and Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998, Mar;8(3):186-94.

78. Kent WJ. BLAT---The BLAST-Like Alignment Tool. Genome Research. 2002, Mar 20;12(4):656-664.

79. Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution. 2013;30(12):2725-2729.

80. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. 2. Cambridge University Press; 2009.

81. Telan EFO, Samonte GMJ, Abellanosa-Tac-An IP, Alesna ET, Leaño PSA, Emphasis YEE, et al. The early phase of an HIV epidemic in a population exposed previously to HCV in the Philippines. J Med Virol. 2011, Jun;83(6):941-7.

82. Grabowski MK, and Redd AD. Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS. 2014, Mar;9(2):126-33.

83. Bobashev GV, and Zule WA. Modeling the effect of high dead-space syringes on the human immunodeficiency virus (HIV) epidemic among injecting drug users. Addiction. 2010, Aug;105(8):1439-47.

84. Wejnert C, Pham H, Krishna N, Le B, and DiNenno E. Estimating design effect and calculating sample size for respondent-driven sampling studies of injection drug users in the United States. AIDS Behav. 2012, May;16(4):797-806.

85. Goel S, and Salganik MJ. Assessing respondent-driven sampling. Proc Natl Acad Sci U S A. 2010, Apr 13;107(15):6743-7.