EXAMINING THE ROLE OF NETWORK STRUCTURE IN HIV TRNAMISSION AMONG PEOPLE WHO INJECT DRUGS IN THE PHILIPPINES: A TALE OF TWO CITIES

Nalyn Siripong

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2017

Approved by:

Brian W. Pence

Ann Dennis

James Moody

Kimberly A. Powers

Annelies van Rie

ABSTRACT

Nalyn Siripong: Examining the Role of Network Structure in HIV Transmission among People who Inject Drugs in the Philippines: A Tale of Two Cities
(Under the direction of Brian W. Pence)

While HIV growth has slowed globally, new HIV epidemics among people who inject drugs (PWID) continue to emerge [1-4]. These epidemics are marked by periods of unusually rapid growth, where HIV increases from virtually 0% to 20-50% within 3-6 years [5-6]. In concentrated epidemics, HIV epidemics among PWID could precipitate larger-scale heterosexual epidemics [7-10], which further underscores the urgent need to find effective ways to reach and deliver prevention services to PWID.

We studied the emergence of HIV among PWID in two cities in the Philippines. The epidemic began in Cebu City, where HIV prevalence grew from 0.6% in 2009 to 50% in 2010 [11]. Expanded surveillance in neighboring Mandaue City found more limited spread of infection, with HIV prevalence reaching only 3.5% in 2011[12], rising to 38% in 2013 [13]. We used exponential random graph models (ERGMs) to simulate network structures and assess whether differences in network structure could explain variation in HIV prevalence patterns in the two cities. We further analyzed genetic sequencing data to consider the extent of overlap or linkage between networks.

Simulated networks showed distinct differences between the two cities. We found smaller network components in Mandaue than in Cebu (1082 v 2980), which may explain the limited spread of HIV there. We also found that Cebu networks exhibited higher degree (21.5

v 10.8), lower clustering (0.29 v 0.39), and shorter average paths (3.3 v 3.8), all of which would facilitate rapid spread of infection across the network. A phylogenetic tree showed high bootstrap support for a large cluster of HIV infections (N = 172), predominantly from PWID from Cebu and Mandaue (85%), which suggests that HIV infection in the two cities arose from a common source infection.

This work suggests that the emergence of a rapidly growing epidemic among PWID in Cebu City spread to Mandaue City, but network structures initially prevented growth of infection. The fragmented network in Mandaue initially prevented penetration of infection into the connected portion of the network. Future work applying transmission modeling to these networks may offer important insights to the role of networks on HIV in these populations.

ACKNOWLEDGEMENTS

This is not my work, but a true collaboration in every sense of the word. I could not have the idea (or the data!) were it not for Genesis Samonte, my ‘ate’ and mentor and friend. I could not conduct the sequencing analysis without Ivo Sah Bandar, and I would not understand any of what I was doing without the guidance of Ann Dennis. I could not understand network concepts without Jim Moody, Jeff Smith, and Jake Fisher. I would have understood anything about respondent-driven sampling without Ashton Verdery. I would never appreciate the importance of modeling without the mentorship and teachings of Kim Powers and Annelies van Rie. I never could have completed this document (and graduated) without my adviser, Brian Pence, who . And I never would have kept my sanity without the support of my friends - especially Christine Gray, Marissa Seamans, and Joann Gruber – and my family, the ‘other’ Dr. Siripongs, and my partner, Edward Levicoff.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: SPECIFIC AIMS

Over 11% of the 14 million people who inject drugs (PWID) worldwide are infected with HIV [14]. 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% in less than a year [6]. 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, may 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 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 [11]. In neighboring Mandaue City, HIV prevalence among people who inject drugs (PWID) was only 3.5% in 2011 and grew to 38%

in 2013 [11-13]. PWID in both cities exhibited risky injecting behaviors, with a majority of people sharing drugs and injecting equipment in both cities. Identifying differences in the network structures between the cities may explain the differences in observed HIV spread in these two cities.

We hypothesize that the networks of PWID 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 (RDS) 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 model (ERGM) 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 PWID infections as a single large cluster that is phylogenetically distinct from MSM or other risk groups; but when we examine these

phylogenetic trees at a finer level of detail, we may find that most of the transmission is happening within each city, as opposed to across the two city boundaries.

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.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

Epidemiology of HIV among people who inject drugs

Among the estimated 12-14 million people who inject drugs (PWID) worldwide, 1.6 million (11%) are infected with HIV [14]. HIV infection presents a serious disease burden among PWID populations, with HIV prevalence exceeding 40% in countries across Southeast Asia, including Thailand and Indonesia [15-17]. 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.

A striking and unique feature of PWID epidemics is the rapid HIV growth, from virtually 0% prevalence to a steady-state level of 20-50% HIV prevalence within 3-6 years. This transition has been characterized by an initial period of very high HIV incidence, with rates reaching 12.5 cases per 100 person-years [10]. As an increasing number of injecting drug users became infected, the pool of uninfected and susceptible PWID remaining is depleted. As the pool of susceptible PWID decreases, the incidence rate fall to rates as low as 2.4 cases per 100 person years [18] and settle at a rate equal to the turnover rate in the injecting drug user population. In rare cases, adoption of safer injecting practices among new injectors can stop spread and reduce overall prevalence [19].

What makes a particular environment ripe for an HIV epidemic? And what puts a particular individual at risk of infection? It is well recognized that differences in HIV prevalence often arise from different environments or behaviors [20]. Rapid spread of HIV

infection has been documented in many PWID populations [21], but other communities maintain low or undetectable HIV levels, despite persistent needle sharing [22,23]. We explore whether network structures may offer an explanation for unusual or unexpected patterns of HIV transmission in communities of people who inject drugs.

This dissertation explores the influence of network structures on transmission dynamics, using surveillance data from PWID in two neighboring cities in the Philippines: Cebu and Mandaue. 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 2013 [11]. In Mandaue, early signs suggested limited spread of infection, as HIV prevalence had reached only 3.5% in 2011 [12]. HIV prevalence grew in Mandaue, but the spread of infection was slower and its reach was limited, with just 38% HIV prevalence in 2013 [13]. PWID in both cities exhibited high-risk behaviors, with a majority of people sharing drugs and injecting equipment. Thus, identifying differences in the network structures between the cities and assessing the impacts of network structure on transmission of infection may offer insights to understanding factors that impact the patterns of disease spread in these two cities.

Aim 1: Importance of networks for disease transmission

Two minimal conditions are necessary for HIV transmission to occur as a result of injecting drug use: (1) one PWID must be infected with replicating virus; and (2) the HIV-infected PWID must use injecting equipment (cookers, needles, or drugs) before sharing it with his uninfected partner. Network structures can play an important role in creating these two conditions, as laid out in Figure 1. In this review, we first discuss how changes in network structures may influence disease transmission based on mathematical models and simulations of sexual and social networks. We then consider past work on how PWID

networks influence decisions to share injecting equipment and ultimately how they may influence disease risk.

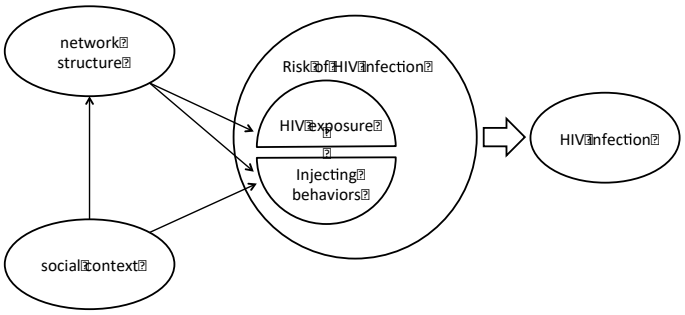

Figure 1. Conceptual diagram describing how networks impact risk of HIV infection.

Understanding the disease dynamics from network theory and modeling

Networks may alter perceived risk of infection in at-risk populations. For example, a person with a single sex partner appears to be at low risk, but may still be connected to a larger sexual network [24]. Similarly, an individual with multiple concurrent partners would typically appear high-risk, but actual risk of infection may be low if each of his or her partners was monogamous. In the absence of empirical data on networks among individuals at risk for HIV, modeling and simulation studies have offered a better understanding of how network structures may affect contact patterns and disease transmission.

Early simulations approximated network structure by modeling differential mixing patterns among people with different contact rates [25,26], following on work identifying the role of core groups in the transmission of sexually transmitted infections [27]. These studies show that homogenous mixing by activity group (i.e., people with many sex partners choose partners who also have many sex partners) would yield faster epidemic spread, but that in the long run, infection would remain more contained than under heterogeneous mixing conditions. However, these models still assume random selection of sexual partners *within*

each group – an assumption that does not typically hold in most empirical sexual or social networks.

Further adaptation of ordinary differential equations (ODE) approximate network effects through two parameters: degree and clustering [28]. Degree is simply the average number of connections a person has. Clustering refers to how frequently closed triangles are observed when three people are connected to each other on the network. Although largely hypothetical, these simulations show how network structures could impact transmission dynamics. Epidemics were less likely to take hold in networks with low degree or high clustering, as both of these measures reduced the effective contact rate. The impacts of clustering were amplified in networks with low degree.

Although increasing average degree was found to be important for promoting disease spread, degree *distribution* can vary widely even while holding average degree constant. Several types of degree distribution and their implications for transmission modeling have been reviewed [29]. In latticed networks, individuals are equally spaced across the network and each person is connected to his or her nearest neighbors, resulting in a relatively constant degree distribution. This network is highly clustered, and risk of disease would be strongly associated with distance from the initial infection. So-called "small-world" networks have degree distributions similar to lattice networks, but their clustered structure is slightly modified so that a few random long-range ties are formed, which creates opportunities for infection to quickly jump and spread across a network in a less rigid manner [29-31].

Some empirical evidence may show that the degree distributions that describe social networks among people have greater variability than the lattice and small-world networks describe above. Data on sexual and social contacts suggested that these networks may

approximate scale-free degree distributions [32-34]. One feature of scale-free networks is that most people have very few (one or two) connections so that the bulk of the distribution is at the low end of the distribution, and only one or two people are in the tail of the distribution, connected to hundreds of others. These high-degree individuals could be considered analogous to the super-spreaders or core groups in epidemics [29,35], and simulations of transmission over scale-free networks often result in epidemics where infection is concentrated around those people with the highest degree [35,36].

Scale-free networks may not be the only relevant degree distribution for understanding transmission of infection. Some argue that this may not be the best distribution to describe these empirical networks [37,38]. Moreover, knowing the degree distribution does not necessarily give us sufficient information to predict how infection would be transmitted over a network. Holding degree distribution constant, other network characteristics can vary considerably, with important consequences for disease dynamics. Increases in clustering, for example, can reduce the growth of the epidemic by creating small inter-connected groups that are only weakly connected to other parts of the network [39]. Average path length on different networks has also explained observed variation in disease dynamics, even after accounting for degree distribution and clustering [40].

Several studies have collected information to explicitly identify ties among individuals on the networks, which further our understanding of how network structures may alter disease dynamics. Adolescents in schools to follow particular "rules" for forming romantic relationships, which result in a sexual network with many connections but limited pathways for infection to spread and consequently highly susceptible to random breaks in the transmission chain [41]. Over time, sexual networks can exhibit greater cohesion through

increased clustering, which would make the network more robust to sustained transmission of infection [42].

Recognizing the wide variation possible even within a constrained set of network parameters (e.g., degree distribution), much work on modeling of disease transmission over networks has adopted a two-part approach, which both estimates the network formation process and simulates transmission of infection over that network. Several studies simulating transmission over sexual networks observed in empirical data found that these networks were susceptible to breaks in transmission [41,43]. Larger networks simulated from casual contact or traffic patterns suggested that more densely connected networks were not subject to the same vulnerabilities [33]. Simulations based on empirical data on social networks estimate the range of potential network realizations for a given set of network characteristics, as well as their consequences for transmission dynamics [44,45].

Pairwise-formation models approximated transmission over agent-based models by considering each tie in the network as a separate observation. While some of these models can reproduce disease dynamics consistent with the agent-based models, they do not record the disease and contact history of individuals on the network [29,46]. These models can incorporate some of the effects of network structures on transmission [47], but they do not reconstruct the complete network, and thus may not be able to identify the roles of larger structures (i.e., those made up of four or more nodes).

Simulation modeling has offered important insights in how network characteristics may translate into increased or decreased risk for different individuals. Analysis of empirical social and sexual networks, combined with the simulation work, helps us understand which

combination of network characteristics has the greatest influence on disease transmission dynamics.

Networks among people who inject drugs (PWID)

Network studies of PWID consider both the effects of local networks on drug use practices and the impacts of network structures on an individual's exposure to infection. Local networks consider an individual's direct social ties and shared bonds or trust can influence his or her risk behaviors. When networks were described more comprehensively, large structures, such as bi-components (cycles), could be identified and their impacts assessed by estimating whether membership to these structures was associated with higher disease prevalence.

Many network studies considered ego-networks, which refer to a central individual, or ego, and all the people directly tied to him or her, also called alters. Such analyses typically used logistic regression to assess how the number and type of ties a person has will impact his or her injecting practices. An injecting drug user who had more fellow PWID as friends in his or her network (i.e., a larger network size) also reported more frequent sharing of drugs and needles [48,49]. The types and strength of ties on a network could also influence the subsequent adoption of injecting behaviors. Injection sharing often occurs in strong social ties, such as family members, sexual partners, or very close friends [50,51]. In addition, cessation of sharing was often most difficult within those pairs with those the strongest social bonds [52]. New injectors often shared equipment with the person who introduced them into drug use [53,54], and the presence of older PWID in one's network could lead to initiation into other risky practices, such as visiting shooting galleries [53,55].

Networks also reinforced social norms that could influence and disseminate injecting practices. HIV-negative PWID shared more frequently with people who they perceived to be HIV-negative [56-58] and in some cases, PWID who knew their own HIV-positive status might refuse to distribute or share their needles with known HIV-negative friends [59]. Such "informed altruism" could prevent spread of infection; but in more recent years, this practice was not maintained [56].

The composition of PWID networks plays an important role in network structure and separation. A strong preference for assortative mixing, or homophily, could result in disconnected groups of people, which might limit transmission of disease across a network. Such preferences often arise from ethnic or racial differences [60,61], but could have consequences for injecting norms adopted among network members [60]. The network consequences of different homophily patterns might partially explain variation in disease prevalence across different groups [62].

Network structure refers to the topology and shape of the network, which defines exposure potential. While network structures could arise in part from dyadic properties like homophily, they also could consider how a network structure of three or more connected people will impact individuals located within or adjacent to that structure. Very few studies collect network data to describe a network's structural properties; in the following review, we describe the major findings from studies that collected broader network data among the PWID population.

Use of PWID network structures to estimate impacts on HIV infection

Colorado Springs

One of the oldest and most widely studied HIV risk networks is the Colorado Springs study, in which researchers collected network data on people at high risk of HIV infection (female sex workers and people who inject drugs, and sexual partners of both groups) and their sex and injecting partners. HIV prevalence among respondents was just 5%, which the authors suggest could be a consequence of limited connectivity in the risk networks [63,64]. Out of the 19 HIV-positive individuals in the sample, only five were part of the connected sexual network component, and four were on the largest connected injecting network component [64]. Most infected individuals were located in isolated groups, which offered little opportunity for transmission. Furthermore, spread of infection might have been limited due to the fragility of the network and low cohesion properties [42,43]. Analysis of follow-up data on the networks of these participants showed that while summary network structures remained stable over time, significant turnover among the individuals who made up those structures could have altered the potential for disease spread over time [65].

New York City

Networks were also studied to understand HIV transmission among PWID in New York City, where HIV seroprevalence reached 40% to 60% [21,66]. PWID were asked to provide extensive demographic information about sexual or injecting partners, which was used to draw the network ties between individuals. One study looked specifically at the 2-core network structure, which describes a subset of the network in which each person is connected to two others on the same network. To classify potential exposure to risk, PWID were classified in one of five groups: (1) members of the largest 2-core (cyclic) structure; (2)

people on the periphery of the 2-core; (3) members of smaller connected components; (4) people who were named by only 1 other person on the network; and (5) people who were completely disconnected from the network. Disease prevalence was highest in the largest connected 2-core (57% HIV and 84% HBV), compared to the other categories, which ranged from 32% to 39% for HIV and 68% to 74% for HBV [66]. The strength of this association remained even after adjusting for effects of the local network, such as composition and size of a person's injecting network [62]. However, the potential for network structures to keep infection contained in pockets or subcomponents of the network, did not explain why the disease prevalence had remained stable in the community for several years [67].

Winnipeg, Manitoba

From 2003-2004, studies of partial network structures among PWID in Winnipeg, Manitoba were used to estimate the potential association between local network structures on disease prevalence. Individual behaviors proved quite important in this population, where disease prevalence (HIV, HBV and HCV) were strongly correlated with injecting with a used syringe and injecting at shooting galleries [68], but network analyses revealed additional important insights. PWID reported low network degree, with an average of about 3 network members, which could explain the relatively low 8% HIV prevalence [52]. In addition, PWID were more likely to share injections with family members or sexual partners than with their friends [52], suggesting that PWID were selective about their sharing partners. This selectivity may have prevented the introduction and widespread transmission of HIV infection in this community.

In 2009, a study of PWID Winnipeg [69] collected demographic information about up to 10 alters in the network, which allowed the authors to draw a more complete picture of the

network in this population. This work showed a sparsely connected network, consistent with the earlier study. However, the reach of this network expanded significantly when ties were imputed among PWID who frequented the same locations [69]. The largest connected component, after accounting for shared geographic space, connected just under half of the 600 PWID in the population. Prevalence of disease in those components was higher than the overall population, and mean path length was quite short (mean distance of 3.6), which suggested an environment that could facilitate rapid disease transmission. Low degree networks previously reported in this population may underestimate the true risk network since they do not account for potential risk exposures arising from visiting shooting galleries or other common geographic locations to shoot up.

Appalachia, Kentucky

A study in rural Appalachia used RDS to recruit PWID and asked participants to name and describe up to 6 injecting network members, who could be matched to provide a picture of a large part of the network in this PWID population. Individual and network characteristics were measured to assess their associations with Hepatitis C, which infected over half (55%) of the population. Individual injecting practices – such as choice of drug and reported needle sharing – were strongly associated with Hepatitis C prevalence. Even after adjusting for these behavioral variables, measures of network centrality were still associated with increased HCV seroprevalence and HCV RNA status [70,71], suggesting that network connectivity played an important role in transmission, above and beyond individual injecting behaviors and practices.

Aim 2: Phylogenetic analysis of HIV infections in Cebu and Mandaue

Phylogenetic methods attempt to describe the genetic diversity of a population by comparing differences in the location and frequency of individual nucleotide bases that arise from random base substitution on the genome. Unlike most pathogens, whose evolution cannot be observed over any period of time less than years or decades, HIV evolves on a time-scale that allows us to apply phylogenetic methods to distinguish between different potential transmission sources [72-74] and identify clusters of transmission among people who share an infection strain [72].

Sequencing data, which are the main data used for phylogenetic analysis, can be combined with epidemiological data on timing of infection or surrounding social networks to further substantiate conclusions about transmission patterns. Phylogenetic patterns of hepatitis C infections in Melbourne, Australia were found to follow the reported injecting networks in the community [75,76]. Identification of transmission clusters in RDS surveys confirmed necessary assumptions of random referral in RDS chains [77,78]. Analyses using timing and geographic locations of detected infections were used to trace the sources of two separate HIV strains in China, identifying and isolating separate points of introduction [79].

In PWID populations, phylogenetic methods have been used to describe potential routes of the HIV spread. Genetic subtyping and other phylogenetic methods were used to trace the spread of HIV infection among drug users to follow drug trafficking routes [79-81]. When new outbreaks emerged, phylogenetic analysis was used to trace their origins to shared risk behaviors [82-84] or geographic location [79-81,85-88]. The emergence of HIV among PWID in Latvia and other parts of Eastern Europe was traced back to ongoing epidemics among PWID in neighboring Russia and Ukraine [85,86]. Epidemics among PWID in Taiwan were linked to HIV infections among PWID in mainland China [87]. In Thailand,

such analysis was used to identify a new recombinant (CRF15_01B) and to trace it back to people exposed to two HIV epidemics – one among PWID and a second primarily among heterosexuals [88].

Phylogenetic methods have also been used to link or isolate different outbreaks of HIV among PWID. When a spike in HIV infections among young PWID was observed in Thailand, phylogenetic methods were used to determine that this new epidemic was distinct from and unrelated to a community older, chronically infected PWID [89,90]. Analysis of a recent HIV outbreak among PWID in Greece identified five circulating HIV subtypes [4], suggesting introduction of infection through at least five independent transmission events and possibly five different communities of PWID that contributed to the ensuing epidemic. In Western Europe, phylogenetic analysis was used to differentiate between epidemics occurring in different regions of Europe by looking at specific mutations in a single HIV subtype [91]. In Australia, two different clusters of Hepatitis C infection were identified, and further analysis suggested injecting networks formed around shared ethnicities [61]. New HIV infections among PWID in Stockholm and Helsinki shared the same HIV subtype, but phylogenetic analysis found two separate sub-clusters of infection that correlated with their geographic locations, suggesting that an single transmission event introduced HIV infection into the PWID population in Stockholm, initiating the HIV outbreak in this community [92].

The work presented here has shown that the use of phylogenetic and network analysis are appropriate to address our aims. Phylogenetic analysis can and should be suitable to determine whether new HIV infections in Mandaue were linked to the ongoing HIV outbreak in Cebu. Several network characteristics have important implications for the connectivity and subsequent spread of infection. This work would build upon the prior research and offer new

insights to understanding the local contexts that describe the HIV prevalence patterns observed among PWID in the Philippines.

Significance

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 has been typically followed by rapid spread, with HIV prevalence jumping from 0% to 50% in as little as 6 months [6]. However, in some communities, the spread of HIV remained limited, despite highly prevalent risky injecting behaviors [20,23]. Understanding the network structure that describes the dynamics of HIV transmission could help elucidate how rapidly and to what extent HIV would spread across a network.

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. In this dissertation, we used sampled data to reconstruct the potential underlying network and assess what role network structures may have played on the disease dynamics in two cities in the Philippines. We leveraged recent advancements in estimation and simulation of networks and the type of network data collected to fit the data as closely as possible. By incorporating network and behavioral data into phylogenetic analysis, we could provide more context and strengthen the interpretability of our results. Triangulation of these methods to describe development of the HIV epidemic could reinforce reliability and robustness of the results.

Innovation

We capitalize on the unique opportunity to identify key differences in network structures that may have explained differences in HIV transmission dynamics between two adjacent cities. The innovation of this study was 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 employed 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 could evaluate the validity of our approach by replicating the analysis in these IDU populations over time.

This was one of the first studies to reconstruct a complete network from respondent-driven sampling data. Analysis of the network and future work on transmission simulations could provide unique information to target, design and deliver prevention interventions to minimize the impact of the HIV epidemic.

CHAPTER 3: METHODS

To investigate the potential role of network structures on HIV growth in Cebu and Mandaue cities, we reconstructed and compared network structures and traced the sources of HIV infection in these two populations. This work comprised two parts. First, we simulated the social networks among PWID in each city and assessed whether differences in network structures were consistent with differences in the HIV prevalence patterns observed (Aim 1). Second, we compared the genetic and spatial distribution of observed infections based on sequencing data (Aim 2). By reconstructing each PWID network, we hoped to better understand the importance and potential impacts of network structure in facilitating the spread and growth of this infectious disease in this population.

Study Population

This study considered networks among people who inject drugs (PWID) in Cebu and Mandaue City in the Philippines. While injecting drug use was not high at the national level, it has been common practice in the greater Metro Cebu, which includes five major cities: Cebu, Mandaue, Lapu-Lapu, Talisay, and Toledo. Most PWID in Metro Cebu inject Nubain (nalbuphine hydrochloride), a narcotic analgesic often prescribed in medical care settings for expectant mothers during labor. Pooling money to purchase drugs was a common practice that facilitates drug-sharing routes. Shooting galleries have been noted as another possible nexus of transmission, as injectors who attended galleries to rent needles were often unaware of the potential disease risk.

Our data were drawn from the 2013 IHBSS surveillance round, which was conducted in Cebu City and Mandaue City in the Spring and Summer of 2013. The survey employed respondent-driven sampling (RDS) methods to recruit PWID in both cities. RDS is a modified snowball sampling method that can generate unbiased estimates from the population of interest starting from a small, purposive and possibly biased subset of the population. Because the RDS sample need not be initiated from a representative subset of the population, this sampling strategy has often been employed when the sampling frame is unknown, as is the case among people at high HIV risk.

Sampling methods: a brief description of RDS

RDS begins with a small group of "seed" participants who are asked to recruit 2-3 friends from their network. These friends are subsequently asked to recruit 2-3 more friends, and the process is continued until the target sample size is reached. Each wave of recruitment is treated as a single iteration on a Markov chain, such that after a sufficient number of waves, the prevalence of sample characteristics should approach an equilibrium state that approximates the true population prevalence [93,94]. Here, we provide an example of how the method might work in two categories of age group, to illustrate the process.

Let us assume that the true population is split in to two categories: group A and group B, with two-thirds of the true population in group A and the remaining one-third in group B. People in group A also have 75% of their friends also in group A. Among members of group B, half of their friends are in group A and the remaining half is in group B.

Starting with four seeds who are all from A, and we ask them to each recruit 2 other people from their network. Given the network compositions described above, and assuming that people select randomly from the relevant universe of network friends, the four seeds

would nominate six more members of group A (75% of the 8 total recruits) and two members of group B (25% of 8). In the next wave of recruitment, those six A's would recruit nine more A's (0.75*12) and three B's, while the two B's would recruit a total of two A's and two B's (each B recruits half A's and half B's). If we exclude the seed wave (which is convention in RDS), our revised estimate of the prevalence of group B in the population after two waves of recruitment is 30%. By the seventh wave, our estimated prevalence of B matches the true prevalence to three significant figures (Table 1). While this example started with a seed population composed of only people in group A, we would have reached the estimates of comparable accuracy and efficiency if we had started with only people in group B (Figure 2).

Table 1. Over several waves of infection, characteristics in an RDS sample will approach the true population characteristics.

Wave Number							
Wave 0	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
4	6	11	22	44	88	176	352
0	2	5	11	22	44	88	176
4	8	16	33	66	132	264	528
Running proportion (excluding wave 0)							
--	0.750	0.708	0.684	0.675	0.671	0.669	0.668
--	0.250	0.292	0.316	0.325	0.329	0.331	0.332

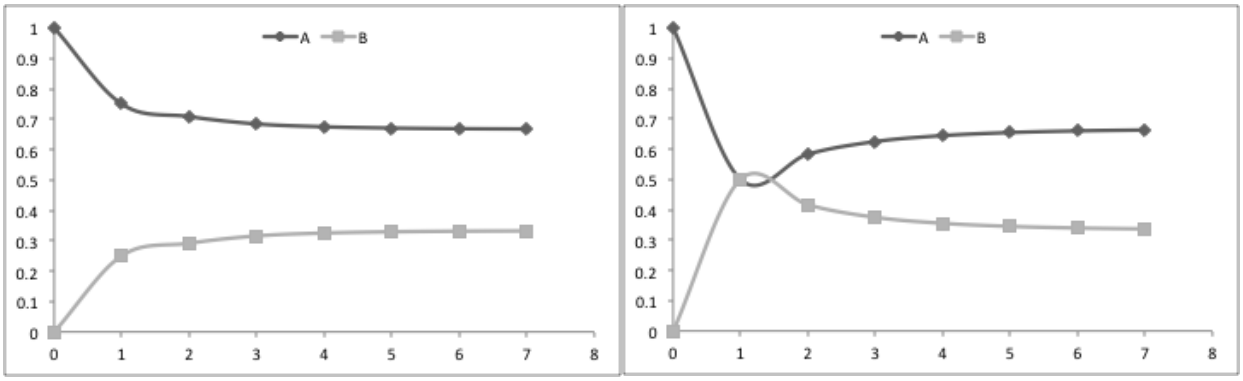

Figure 2. RDS samples will approach true population estimates, whether our seeds are all in group A (left) or group B (right).

This method is subject to a number of assumptions:

(1) The network is completely connected (i.e., we can start from any person and reach any other person on the network within a finite number of steps);
(2) Seeds do not have to be representative of the population, but they should be independently sampled;
(3) We have non-branching recruitment chains;
(4) All ties are reciprocated: this means if X recruited his friend Y, Y would also name X as a friend in the same network;
(5) Degree is accurately reported;
(6) Random referral: people randomly recruit friends in their network (everyone in a person's network has an equal probability of being recruited);
(7) Samples go through sufficient number of waves to reach equilibrium; and
(8) Sampling with replacement.

If all these assumptions hold, RDS can generate unbiased estimates of population characteristics from a non-probability-based sample of seeds. We relaxed assumptions 3 and 8 by using the successive sampling estimator [95], and several other assumptions could be verified in the data. Several additional diagnostic tests were also conducted to assess whether sampling chains reached equilibrium [96].

Most RDS studies are designed to estimate HIV prevalence or population size and ignore the inherent network properties captured by this method. Since RDS uses the peer network to identify recruits and asks about network size, these data can be used to describe the social networks of these populations.

RDS implementation

The 2013 IHBSS round for PWID in Cebu and Mandaue used RDS to recruit PWID in both cities. Seven seeds were identified in each of the two cities to initiate recruitment chains. Peer educators, who were part of the previous Global Fund HIV prevention programs, identified PWID from a range of age groups and geographic areas within each city as "seeds." The RDS recruitment system was described to the seeds before the start of data collection, and they were encouraged to bring or inform people they would choose as recruits to go to the study site, to expedite the recruitment process.

A person who presented a coupon was eligible if they had not previously participated in the survey and also fulfilled three other eligibility criteria: (1) they had to be over 15 years old; (2) living in Cebu province; and (3) reported injecting a drug that was not prescribed to them in the last six months.

All eligible participants completed an interview about demographic and behavioral characteristics. This was followed by a blood draw to test for syphilis, Hepatitis C (HCV) and HIV. Participants received 200 Philippine Pesos (PHP, equivalent to just under US$5) as a transportation allowance and were given two coupons, which they were asked to give to friends who were willing to participate in the survey. To further promote recruitment, people were offered a secondary incentive if they were able to successfully recruit new people in the survey. They were offered 50 PHP (just over $1) per successful recruit, with a maximum of 2 recruits per person.

When recruiters came to the study site to collect their secondary incentives, a very brief, secondary interview was administered, which asked about three areas:

(1) How many people (if any) refused the coupon, which revealed how much of his network was saturated (i.e., people who refused because they had already participated, people who refused because they were not interested).

(2) Reciprocation and characterizing his tie with his recruit; in the primary interview, we asked the participant to describe his relationship with his recruiter, so in this interview we asked the question of the recruiter himself, to confirm whether the descriptions of the ties are reciprocated.

(3) Potential social links between his recruiter and his recruit. Because of the anonymous nature of RDS, each person explicitly knows (at most) 3 people in the study: the person who recruited him, and the two people he recruited into the study. But we can ask this person about the ties between his recruiter and recruits, which allows us to identify up to 2 additional ties on the network.

Recruitment took place from May 27 to June 20, 2013 in Cebu and May 29 to July 8, 2013 in Mandaue. The target sample sizes of 450 from Cebu and 300 from Mandaue were determined based on the estimated population size and availability of funds to conduct the survey. Seven seed PWID from Cebu City recruited 450 other PWID into the survey, for a total of 457 people. Another seven seeds from Mandaue recruited 303 other eligible friends, for a total of 310 PWID in this survey.

Aim 1 Analytic Methods

Estimating population characteristics

To estimate population characteristics, we weighted 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. This is analogous to a Thompson-Horwitz estimator, used in typical in survey sampling. A number of adjustments or revisions have been added to estimate unbiased RDS-estimates under relaxed assumptions [95,97,98]. We employed one of the most recent of these methods: the successive sampling estimator [95], which did not require assumptions about non-branching chains and sampling with replacement. This method would proceed through an iterative process to converge on unbiased population-level estimates.

RDS Homophily

In this analysis, population homophily was expressed as the proportion of extra, or additional, within-group ties, compared to what would be expected if mixing occurred at random, holding degree across groups constant. To illustrate this concept, let us again consider the two groups considered earlier, where group A was 67% of the true population and the remaining 33% was of group B. Let us assume that group A has an average degree of 6 and group B has an average degree of 12. Given the estimated proportions of people in each group and assuming a network size of 1000, we would expect the number of in-group and out-group ties under random mixing assumptions to follow those given in Table 2.

Table 2. Homophily is calculated based on the expected ties under random mixing (left) compared to observed mixing (right).

	A	B	Total
A	2,670	2,666	5,336
B	2,666	2,662	5,328
	5,336	5,328	10,664

	A	B	Total
A	3,286	2,050	5,336
B	2,050	3,278	5,328
	5,336	5,328	10,664

Homophily = (3286 + 3278) / (2670 + 2662) = 1.23

Homophily was calculated as the ratio of the empirical number of in-group ties, compared to the expected number of in-group ties, holding the degree of the two groups constant. In this case, the number of same-group ties (A to A and B to B) was 23% higher than what we expected at random and homophily would be calculated as 1.23. The measure of homophily was centered on 1 (completely random mixing) and represent the proportion of in-group ties more (homophily > 1) or less (homophily < 1) than expected at random.

To estimate homophily from the RDS sample, we used the RDS-weighted degree for each of the groups, to reconstruct a table of in-group and out-group ties expected with random mixing. We then estimated the number of in-group and out-group ties in our population, based on recruitment patterns, weighted by individuals in the network. In this analysis, we considered homophily on variables of where and how people inject.

Measuring the ego-network configuration distribution

The ego-network configuration was used to fit the clustering network parameter in our network. Ego-networks included data on a central person (ego) and all the people directly connected to him (alters) and all the ties among them. The ego-network configurations considered ties among the alters in the ego-network, *excluding* their ties with the ego.

In our RDS data, each person in the recruitment chain was also the center of an ego-network of 3 other people: his recruiter (i.e., the person who gave him the RDS coupon) and

his recruits (i.e., the people to whom he gave RDS coupons). We could not directly ask whether the recruiter (person X) knew his recruit's recruits (Y and Z). However, the central ego (U) knew that all three of these people were in the survey, and so U can provide information on whether these alter-alter ties (between X, Y, and Z) exist.

Data on the recruiter-recruit ties (X to Y, X to Z) were collected at the follow-up interview, when individuals returned to receive their secondary incentive. The survey did not explicitly ask if the two recruits knew each other; so data on the Y to Z tie was imputed based on the proportion of dyads that had ties in each city. This meant only three of the four possible ego-network configurations could be observed (first three figures in Figure 3) and the fourth could occur if an additional tie was imputed. We then weighted these configurations using RDS-weights for each ego in the network.

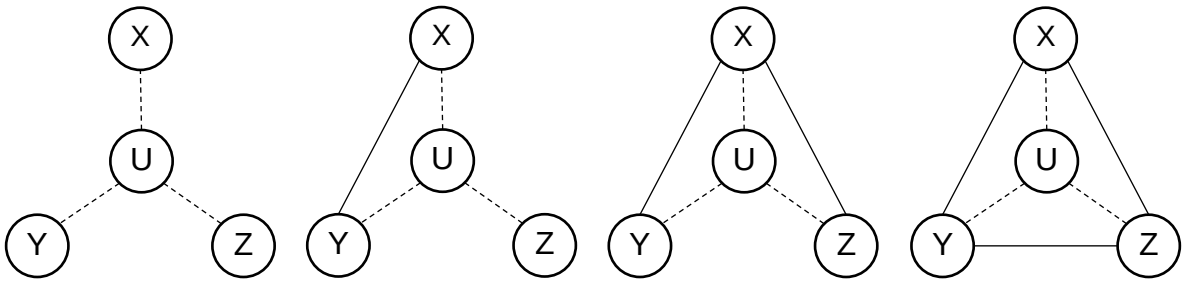

Figure 3. Ego-network configurations observed from RDS data collected in our survey.

Simulating networks

To estimate the underlying networks in these two cities, we used a method that estimates parameters for an exponential random graph model (ERGM) by simulating networks and adjusting model parameters that were consistent with the observed networks in our two cities, as demonstrated by Smith [99] and Merli [100]. The method used ERGMs in an iterative process that started from an initial guess of model parameters, which were

subsequently adjusted until model fit could not be further improved. The parameters of interest in our model included mixing parameters, which aligned with the homophily in the population; and clustering ties, which we fit to the observed ego-network configurations in our empirical data.

Exponential random graph models

Exponential random graph models (ERGMs) are statistical models that estimate how network properties, such as homophily and transitivity, impact the formation of ties on a network. They follow the form:

$$\Pr(X = x) = \frac{\exp\{\theta' z(x)\}}{\kappa(\theta)}$$

where X is a realization of the network structures, $z(x)$ refer to the network statistics (homophily, clustering, etc.), and $\kappa(\theta)$ is a normalizing constant, so that probabilities sum to 1 [101-104]. In all but the simplest of cases, the normalizing constant expressed in this equation is too computationally complex to calculate. We avoid estimating this constant by estimating the conditional probability of a single tie, holding the rest of the network constant. Note that if we compare the odds of the presence of a tie vs. the absence of a tie, conditional on the rest of the network, our estimation equation simplifies to:

$$\frac{p\left(X_{ij} = 1 \middle| X_{ij}^{c}\right)}{p\left(X_{ij} = 0 \middle| X_{ij}^{c}\right)} = \exp\left\{\theta'\left[z\left(x_{ij}^{+}\right) - z\left(x_{ij}^{-}\right)\right]\right\}$$

The coefficients actually represent the effects of a unit change in network statistics on the odds of a tie in the network. If θ' on a same-group mixing parameter (e.g., A-A mixing) was positive in sign, we would expect to see that two people of group A were tied more frequently than expected at random. These models have been used in sociology to estimate

associations that describe how adolescents choose friends based on a combination of age, gender and race [102]. It has also been used to simulate networks when partial network data were available [99,105].

Fitting the ERGM formula

We seeded these models with two forms of data: (1) an initial network with the joint distribution of characteristics, including degree, of the population; and (2) an initial set of mixing parameters, drawn from the data on observed ties.

Set initial parameters: the initial network

Before we could estimate the network structures, we first estimated characteristics and properties of individuals in the network. We used the RDS weights to estimate the joint distribution of demographic and behavioral characteristics in the populations in each city, and applied this to the estimated PWID population sizes in each (3000 PWID in Cebu, 1500 PWID in Mandaue). The characteristics of interest included: age, categorized into three groups; age at first injection drug use; visits to shooting galleries and tambayan (private injecting parties); frequency of injection, drug of choice, and source for needles and syringes.

To apply degree (number of ties) of individuals in the network, we drew degree from a smoothed log-normal degree distribution, for our 3000 and 1500 PWID populations, respectively. These draws were then matched to individuals with appropriate characteristics, so that the characteristics by degree rank and average degree across groups were consistent with the RDS-weighted results. Once all characteristics were assigned to the two populations, we used an algorithm that creates ties with people on the network, based on an assigned degree [106].

Initial homophily

Next, we provided ERGM mixing parameters of homophily. In the ERGM formula, these mixing parameters could be interpreted as "the odds that two people of the same group are tied." We could estimate an analogous logistic regression parameter if we considered the unit of analysis to be a dyad (that is, any pair of two individuals) and the outcome of interest was the presence or absence of a tie. If we collected data on all the observed ties (recruitment ties in the RDS survey) and the homophily among them, compared to a random sample of observed non-ties (a random sample of dyads from the RDS survey), then the resulting logistic regression would estimate the odds of a tie, given the possible characteristics of dyads (A-A, B-B or A-B, for example). The resulting coefficient estimate values were then applied as an initial starting point for the analogous coefficients for mixing, in the ERGM equation.

Initial clustering (GWESP)

Clustering measured the potential of two people having a mutual friend, thus creating a closed triangle on the network. In an ERGM, the clustering parameter could be interpreted as the odds of a tie between two people with the same mutual friend. Unlike measures of homophily, clustering could be estimated in a dyadic logistic regression model, because it is dependent on the presence or absence of other ties on the network, which was the "outcome" of our logistic regression model for homophily. The dependent structure of the clustering coefficient made it quite difficult to estimate because of its vulnerability to model degeneracy, a consequence of the fact that the closure of one triangle could create many more open triangles. Instead, we parameterized the clustering coefficient by using the geometrically-weighted edge-wise shared partner index (GWESP), expressed as:

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

where α set the rate of decay for each additional partner and p_i was the number of connected dyads who had individual i in common. We initially set our GWESP parameter to 1, and allowed the fitting process to estimate the correct value.

Network estimation process

Using mixing parameters set according to observed homophily and starting the clustering (GWESP) parameter, we then sampled coefficient values within a range (+/- 1) of our original starting point, and then simulated 200 networks for each set of sampled parameter coefficients. To reduce the sample space of possible networks, we constrained our fits to only those networks with the same degree distribution as the initial random network.

Based on these simulated networks, we extracted the networks that most closely match the ego-network distribution, based on the chi-square statistic. We then reassessed homophily coefficients on those by comparing ties in empirical data (cases) to ties on the simulated networks (controls) to estimate the difference between simulated and observed homophily. These estimates were then used to adjust the corresponding ERGM coefficients to account for this difference. Using these updated ERGM coefficients, we simulated a second set of networks while varying the clustering parameter. We then compared the ego-network distributions on this new set of simulated networks to the empirical distribution, to identify the best-fitting clustering parameter. Updating ERGM parameter for clustering may have changed homophily effects, so these parameters were reassessed and fitted. This process was repeated until we arrived at ERGM formula with the best-fitting ego-network distribution. This best-fitting formula was then used to simulate 50 networks in each city.

Using this set of 50 networks, we compared network properties, such as homophily statistics, with RDS estimates, to check that the networks we simulated were consistent with the observed data. Consistent networks did not guarantee that our model fully described the true underlying network, but they offered one plausible version of the networks we observed. We then analyzed the properties and characteristics of these simulated networks, in an effort to identify potentially important individuals on the network using centrality measures.

Network Analysis

We analyzed simulated networks to describe the global network characteristics and to identify potentially important individuals within the networks. We looked at several characteristics relating to the connected components and centrality on the networks, as described below.

Network components

A network component captures all individuals who can reach each other through a finite number of steps or ties [107,108]. Three components are depicted in Figure 4. Component A is the largest connected component and includes the largest proportion of people in the overall network of 12 individuals. Assuming ties do not change over time, the largest connected component of a network defines the true maximum reachability of a single infection across the network. For each city's set of networks (Cebu and Mandaue), we estimated the number of components, their average size, and the size of the largest connected component on each network.

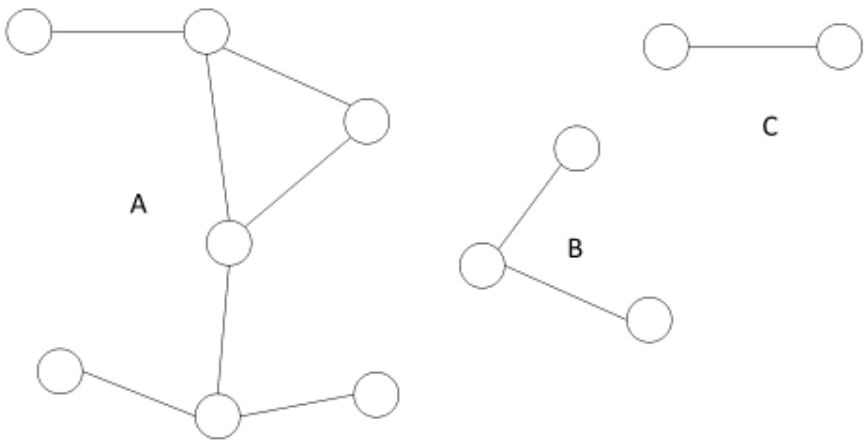

Figure 4. Three components make up this network of 12 people; component A is the largest connected component, including 58% (7 out of 12 individuals) of the network.

Geodesic path length and average distance

Within the largest component, we measured the geodesic, or the minimum number of steps it takes for each person to reach every other person on the network [107,108]. Geodesics were measured for every pair of individuals on the same component, so that for a component of size *n*, we estimated $n * (n-1) / 2$ paths. People who were not on the same component were an infinite number of steps away, so their geodesic could not be computed. We calculated the average distance, or the average of all geodesics, on the largest component in each of the two cities.

Network centrality

Network features such as membership to a 2-cores [67] or occupation of particular positions in the network may increase or reduce risk of infection for different people. To identify some of these positions, we estimated four common measures of centrality: degree, eigenvector, closeness, and betweenness centrality [107-109].

Degree: The most basic measure of how much a network influences a person is how many ties a person has, which is called their degree. A person with high degree centrality is well connected and thus has a higher potential for exposure to infection. Increasing degree is therefore highly correlated with higher risk of infection. Similarly, people with higher degree (more connections) can reach a large number of people within a few steps and are also at high risk for onward transmission.

Eigenvector: Degree does not always fully capture centrality. A person with low degree can be connected to one or more persons with very high degree, so that their low degree does not necessarily reflect their reach or connectedness along the network. Eigenvector centrality captures these higher-order effects of being one or more steps away from very high degree individuals along the network. It is measured as:

$$C_E(n_i) = \frac{1}{\lambda}\sum_k a_{k,i} n_i$$

where $a_{k,i}$ is the adjacency matrix of which nodes are tied; and λ is a constant.

Closeness: This measure reflects reachability and relative distance between people on the network. A person with high closeness centrality can reach a large number of others in a small number of steps. For an individual person on the network, closeness is measured as the average of the inverse-distance he is from every other person on the network:

$$C_c(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1}$$

where the function *d* is the distance function, measuring the shortest path between nodes n_i and n_j. Individuals with the highest closeness centrality can quickly reach the largest number of people within just a few steps.

Betweenness: Individuals with high betweenness centrality lie on many of the shortest paths connecting other people in the network. The measure is a count of the total number of shortest paths an individual lies on.

$$C_B(n_i) = \sum_{j<k} g_{jk}(n_i) / g_{jk}$$

where g_{jk} is the number of shortest paths between *j* and *k*, and $g_{jk}(n_i)$ is the number of shortest paths between *j* and *k* that go between node n_i. Individuals with high betweenness centrality may be gatekeepers or important bridges between these individuals. However, if there are many paths to between the two individuals, then high betweenness (lying on many of the *shortest* paths between others) may be of less importance.

Centrality measures tend to be highly correlated, but poor correlation between centrality measures may signal unusual or important structures and particularly unusual individuals on the network. For example, individuals with high betweenness but low degree centrality indicate people who may serve as important bridges between small or large groups of people.

Clustering coefficient

To measure clustering on the network, we use the clustering coefficient, which is a measure of the number of observed triangles on the network as a proportion of the number of possible triangles:

$$\phi = \frac{\text{3 x number of triangles on the network}}{\text{number of connected triples of vertices}}$$

We consider here at the effects of clustering on component A from Figure 4 as an example. In this component, there are 10 connected triples and one triangle, for a clustering coefficient of 3/10 = 0.3. The clustering coefficient does not scale in a formulaic or

predictable manner with the addition of a tie that will close the triangle, as the closure of one triangle may create a number of new potential triads that were not previously present on the network. In Figure 5, we can add one additional tie to the original component A in two different ways that could result in different measures of the clustering coefficient.

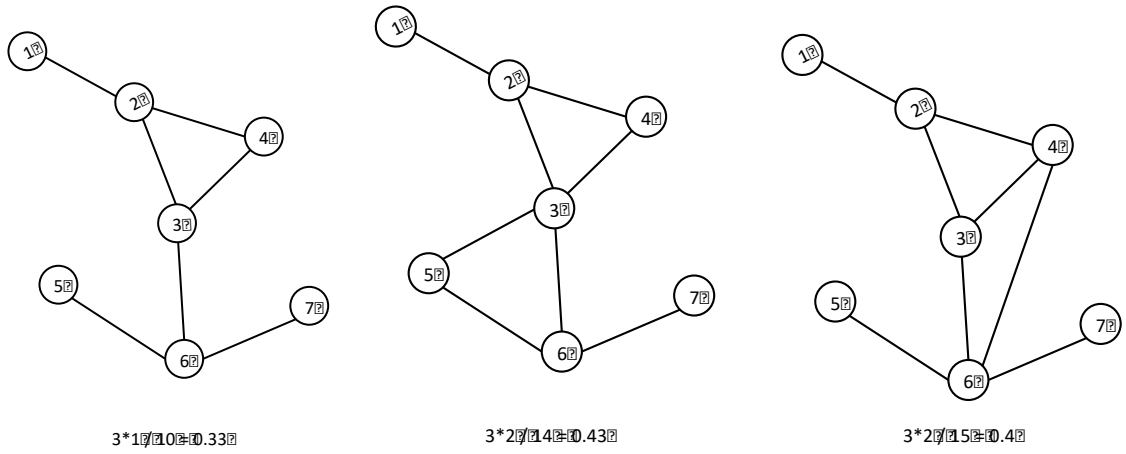

Figure 5. Adding one tie to a network changes the clustering coefficient, but in unpredictable ways.

Clustering can have two consequences for network structure. When clustering is very low, increased clustering can create greater network cohesion, which creates more available paths for infection to travel. Low network cohesion could make the spread of infection may more susceptible to random breaks in transmission that might occur with diseases with low per-contact transmission probabilities [41]. Therefore, some degree of clustering might be needed to ensure the network is sufficiently connected to sustain spread of infection. Beyond a certain level of cohesion, however, added clustering results in the addition of paths that may be redundant because the people they connect are likely to already be infected through other pathways.

We then considered whether differences in these network characteristics aligned with our expectations, based on the observed differences in the timing and HIV prevalence patterns in each city.

Aim 1 Limitations

This work employs methods that remain quite new in the area of network estimation and simulation. It may provide potential insights to network structure, but our results are limited in the data we have. Three major limitations were:

(1) Reliability of RDS data: Although respondent-driven sampling is employed widely, there remains some controversy over whether estimates or RDS-weights and uncertainty are accurate [110]. The demographic composition of network simulations, then, were only as reliable as our RDS estimates. Given the difficulty in recruiting this population, however, it was the best available data in this population.

(2) Truncated ego-network distribution: Fitting of the clustering parameter was based on our ego-network distribution. Given the format of our data, our model fit was compared to the distribution of only four possible ego-network configurations, which meant that there could have been a wide range of possible parameters that matched this distribution, and thus we could not guarantee that our coefficient estimates were the singular best solution.

(3) Fitting process was limited: The network estimation process was not comprehensive, as the space for parameter estimation was intractably large. We constrained the degree distribution to limit one dimension of the parameter space, but a large number of other network configurations were possible.

Aim 2 Analytic methods

The use of phylogenetic analysis is particularly useful in the study of viral evolution because of the high mutation rate that is unique to viruses. Unlike most organisms in the world, HIV does not have the DNA polymerase and its reverse transcriptase makes multiple base substitutions in every copy of genetic material [111]. We can analyze the genetic material of viruses to identify different generations of the virus to reconstruct a sort of genetic 'family tree'.

Building the consensus sequences

Each successfully sequenced sample had a large set of 50,000 to 100,000 'clean' reads. To generate the phylogenetic tree, we first had to reduce each set of data to a single consensus sequence that reported the majority nucleotide base (A, C, T, or G) present at each position in the sequence (Table 3). In positions with no clear majority (or a plurality), a placeholder (X) was used to indicate that a base was reported but no clear majority was observed in the data. In the alignment, these were treated as deletions, or missing data. The Center for AIDS Research lab at the University of Hawaii at Manoa used in-house software, Intergroomer (http://courge.ics.hawaii.edu/inte/groomer/), to generate consensus sequences from the data.

Table 3. Building a consensus sequence usually means taking the majority base present at each site.

DATA									
1	A	A	A	C	T	G	A	G	G
2	A	A	A	C	T	G	T	G	G
3	A	A	G	C	T	C	A	G	G
4	A	A	A	C	T	C	T	G	G
5	A	A	G	C	T	G	A	G	A
CONSENSUS	A	A	A	C	T	G	A	G	G

To properly construct the phylogenetic tree, we first added 286 reference sequences, which we obtained by searching GenBank for all HIV sequences from the Philippines and the 10 closest matches with our sequence data. These data were combined into a single alignment file.

Sequence alignment

The consensus sequences were aligned with the HXB2-LAV K03455 reference genome using MUSCLE [112,113]. The alignment process was needed to account for errors in the replication process that could result in random insertion or deletion of a base. If we had not properly account for these insertions and deletions, we would overestimate the degree of genetic diversity because some sequences would have been incorrectly shifted by one or two positions (Figure 6). To correct for this, the MUSCLE algorithm iterated through a number of steps to arrive at the best alignment.

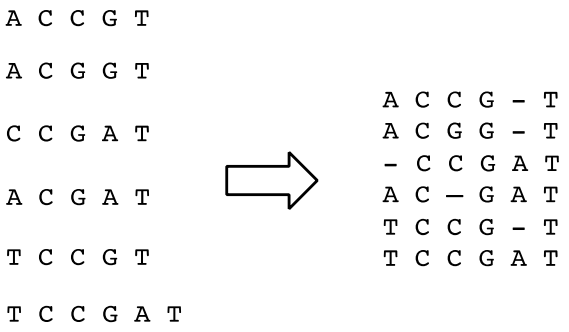

Figure 6. Insertion of blank spaces can be necessary to align sequences with each other.

The MUSCLE algorithm followed several steps to properly align sequences. First, genetic distance was measured and a basic phylogenetic tree was constructed to cluster sequences based on the data provided, assuming no inserted or deleted positions. Next, sequences were aligned within each cluster, making appropriate insertions as needed for an

initial alignment (MSA1). Using these sequences (with the new insertions), genetic distances were recalculated and another tree was generated, resulting in a second alignment dataset (MSA2). This new alignment (MSA2) was assessed by splitting the new tree into two subtrees. The two subtrees are aligned separately and then rejoined, inserting gaps where needed. The two alignments were scored based on the number and size of gaps added in each sequence; if the score improved, the new alignment was again reassessed. Otherwise, the old alignment was used. Alignments were further confirmed by visual inspection, to ensure that none of the results arose from an unexpected artifact of the data [112,113].

Phylogenetic tree construction

Once our sequences were aligned, we used RAxML to find the best fitting maximum-likelihood phylogenetic tree [114]. The maximum-likelihood method fitted a model of evolution and tree topology to the sequence data to estimate the best-fitting tree. Specifically, we maximized the likelihood function such that the tree τ and evolutionary model M generated the highest probability of the data, which consisted of base substitutions occurring at site j for sequences of length l across all sequences in the alignment (D).

$$L(\tau, M, \rho|D) = \prod_{j=1}^{l} \Pr\left[D_j, \tau, M, \rho_j = 1\right]$$

We assumed a generalized time-reversible (GTR) evolutionary model, in which nucleotide substitution rates were defined using the Q matrix:

$$Q = \begin{array}{c} \\ \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cccc} \text{A} & \text{C} & \text{G} & \text{T} \end{array} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & a\mu\pi_C & b\mu\pi_G & c\mu\pi_T \\ a\mu\pi_A & -\mu(a\pi_A + d\pi_G + e\pi_T) & d\mu\pi_G & e\mu\pi_T \\ b\mu\pi_A & d\mu\pi_C & -\mu(b\pi_A + d\pi_C + e\pi_T) & f\mu\pi_T \\ c\mu\pi_A & e\mu\pi_C & f\mu\pi_G & -\mu(c\pi_A + e\pi_C + f\pi_G) \end{pmatrix} \end{array}$$

In the above matrix, π_s represents the observed frequency proportion of base s and thus sum to 1. The letters *a-f* are multipliers of the rate of substitution (μ) for each base. We can see the "reversible" aspect of the model in that the rates of change between the different bases are equal, as evidenced by the fact that the letters are symmetric across the diagonal of this matrix. For example, the substitution rate in of A for C (shown in the first column of the second row) is equal to the substitution rate of C for A (second column, first row).

These methods simultaneously estimated the phylogenetic tree and the model parameters listed above, so they invoked heuristics that allowed us to fit each sequence in the tree so that the tree and model best described the pattern of nucleotide substitutions that we observed. If we had added each sequence in a step-wise manner, we would guarantee reaching a local maximum of the likelihood function. Therefore, we exchanged random branches to rearrange the tree and refitted the model to improve robustness of the model.

To assess branch support or the reliability of our tree branching patterns, we conducted bootstrapping, in which columns (specific sites) of all the sequences were selected with replacement and then used to construct phylogenetic trees using the maximum-likelihood process described above. This sampling and estimation process was repeated multiple times. We presented a consensus tree with bootstrap values that represented the proportion of these bootstrapped samples in which each specific branch was observed.

RAxML employed randomized stepwise addition parsimony trees as the starting point for fitting the model, and then randomly rearranged portions of the tree to search for better model fit [114]. We assumed a generalized time-reversible model, which estimated all evolution parameters from our data [58]. We ran 20 maximum likelihood searches plus 1000 bootstrap replicates.

No well-defined thresholds exist to determine or identify transmission clusters. Prior work had used somewhat arbitrary values for minimum genetic distance and high bootstrap support [115,116]. Although genetic distance may have been confounded by time as infection continues to evolve within given individuals, this was unlikely in our data, given the very recent emergence of the epidemic in Cebu. In our analysis, we wanted to assess whether we can observe any separation was evident between HIV infections from PWID in Cebu and those in Mandaue.

Aim 2 Limitations

One challenge in the analysis was the low success rate of viral amplification, which resulted in a smaller sample size. Blood and plasma were 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% was quite low, there was no indications that degradation was differential with respect to city.

Since the data were collected primarily for public health purposes and not research, the sequencing data were targeted to look in-depth along those regions of the genome that carry clinically important mutations. As a result, the relatively short length of the consensus sequences may not have been sufficient phylogenetic signal to detect underlying differences between two or more different patient sequences.

Another limitation was that our data are missing information about treatment and duration of infection. Antiretroviral treatment exhibits selective pressures on the virus, so that people on treatment would have greater genetic variation, as the virus would attempt to evade

the effects of treatment [111]. While treatment status would be a serious concern in most phylogenetic studies, hospital records (as reported at the central level) indicated that no more than 30 PWID had initiated ART at the time of the 2013 surveillance, so convergent evolution was not expected to be a significant concern in this population.

CHAPTER 4: IDENTIFYING NETWORK STRUCTURES THAT IMPACT HIV TRANSMISSION DISEASE DYNAMICS IN THE PHILIPPINES

Introduction

While the global HIV response has slowed the epidemic in most of the world, an estimated 2 million new HIV infections occur every year [117]. In the Philippines, only 3500 cases were reported in the first 25 years since the first case was reported in 1984 [118,119]. Since 2008, however, the HIV epidemic has grown at a rapid pace, with the number of reported HIV and AIDS cases increasing by 20-40% each year and over 80% of the 32,000 HIV and AIDS cases reported since 1984 occurring within the last 5 years [119]. In a world where HIV is largely on the decline, the HIV epidemic is just beginning to take hold in the Philippines.

People who inject drugs (PWID) are a highly affected population in the HIV epidemic in the Philippines, making up less than 1% of the total population but over 10% of reported HIV and AIDS cases in the country [119,120]. HIV epidemics among PWID can grow rapidly and without warning, with HIV prevalence rising from virtually 0% to 50% in as little as 6 months in India [6] and incidence rates rapidly reaching 12.5 cases per 100 person-years in Thailand [10]. Moreover, it is often difficult to predict when an epidemic will emerge. In the Philippines, widespread needle-sharing behaviors among PWID were noted as early as 2000, but from 1984 to 2009 only 9 reported HIV cases were transmitted through shared needle injection [118,119], with no warning of the rapid rise in HIV that followed.

Most new HIV infections among PWID in the Philippines occurred in Cebu, the country's second largest metropolis. In 2009, HIV surveillance in the metropolis center of Cebu City captured the emergence of an outbreak among people who inject drugs (PWID), where HIV prevalence rose from 0.6% to 53% in the course of one year [11]. Expanded surveillance in neighboring Mandaue City suggested that infection spread more slowly and reached a smaller proportion of the population, even though PWID in both communities exhibited risky injecting practices that would suggest environments ripe for rapid HIV transmission. We investigate whether the network structures that connect PWID within these two cities may explain some of the variation in transmission dynamics observed in the two communities.

Methods

Study population

In 2009, routine surveillance in Cebu City, the Philippines, captured the emergence of an HIV outbreak among PWID. Though only 3 HIV infections were among sampled PWID, surveillance captured a number of additional HIV infections among female sex workers that were also linked to injecting drug use, leading to concerns of an outbreak of HIV among PWID in Cebu. A rapid assessment in 2010 confirmed that HIV was expanding among PWID in this community, with over 50% of PWID infected with HIV. The following round of HIV surveillance in 2011 expanded to the neighboring Mandaue City, part of the greater Cebu Metropolis Region. While HIV prevalence in Cebu City remained high (53%), HIV had not yet spread significantly in Mandaue. Using network-related questions implemented in the 2013 round of surveillance, we sought to estimate the network structures in Cebu and

Mandaue cities, and explore whether differences in these structures are associated with differences in the transmission dynamics of HIV in the two cities.

Network parameters of interest

Network structures and characteristics can influence the potential spread of disease in several important ways. We describe both network-level and individual-level characteristics to assess their potential effects on transmission dynamics. At the network level, we identify all network components. A network component is the set of people who can reach each other through network ties. If disease can only be transmitted through these ties, the component defines the potential number of people who could be infected by the introduction of disease on the network or the boundary of infection spread. The size of the largest connected component should help us determine the maximum final size of an epidemic on a network. We measure what proportion of the network is captured in the largest component to define the potential reach, or maximum prevalence of infection. We also estimate average distance, which summarizes the shortest paths (geodesics) between every pair of PWID on the component. Shorter average distance will accelerate the speed of spread, because we could reach more of the network component in a fewer number of steps. Density of the network is the ties observed on the network, expressed as a proportion of all possible ties. This measure may also impact the connectivity as increasing network density may provide the opportunity shorten paths on the network. We finally measure a clustering coefficient, which is a measure of the proportion of closed triangles on the network. In general, we would expect increases in clustering to create redundant paths on the network and slow transmission [28,39,40].

At the individual level, we compared several measures of centrality, which may identify people who may be important to the reach and connectivity of the network. Degree is

a measure of the number of direct ties or contacts an individual has on the network. Higher average degree on a network should increase the probability of an epidemic occurring [28]. Eigenvector centrality combines higher-order effects of degree (i.e., being connected to someone of high degree will increase eigenvector centrality) and has been associated with risk of infection among PWID [70]. Closeness measures the inverse-path length between an individual and all others on the connected component. Betweenness counts the number of times an individual lies on the shortest path between two other individuals on the network. These centrality measures may identify important individuals or groups of individuals who may influence transmission dynamics across the network.

Data collection methods

PWID were recruited using respondent-driven sampling (RDS), where participants were asked to recruit their friends whom they knew to fulfill the eligibility criteria. A small group of "seed" PWIDs was selected by local health office staff to participate in the survey. After completing survey and providing blood for HIV, HCV and syphilis testing, they were asked to recruit two of their friends whom they knew to inject drugs. These friends made up the first "wave" of recruitment and were asked to recruit up to two more friends who also injected drugs. This recruitment pattern continued until reaching the target sample sizes of 450 participants in Cebu and 300 participants in Mandaue. Each PWID was provided 200 Philippine Pesos (US$5) for participating in the survey and 50 pesos (US$1.25) for each eligible person they recruited. When people returned to redeem the recruitment incentives, a brief secondary questionnaire was administered to identify ties between the participant's recruiter (the person who gave him his coupon) and his recruits (the people he recruited into the survey).

Statistical analysis and network simulation

Using methods proposed by Smith in 2012 [99] and previously applied to RDS data [100], we estimated coefficient parameters for an exponential random graph model (ERGM), a conditional regression model that relates network characteristics, such as mixing and clustering, with the presence or absence of a network tie [101,103,104].

Our network simulation started from a random network with node characteristics that matched our RDS-weighted population. This random network was combined with an initial set of coefficients for a defined ERGM. A series of new networks were simulated from these coefficients, and their results were adjusted to more closely match the empirical network characteristics of mixing and clustering, as estimated from our RDS-weighted data. The process could be described in four parts. First, we estimated characteristics of individuals on the network and overall network characteristics (mixing and clustering) from the RDS data. Next, we set up the initial ERGM parameters and simulated the networks. Third, we iteratively adjusted ERGM coefficients and simulate new networks until our coefficient estimates closely resembled the empirical data. We then simulated a number of networks from our final set of coefficients. All analyses were conducted in R statistical software [121], using the RDS [122], igraph [123], and statnet suite [124,125] of packages. We discuss each step in further detail here.

Estimates from RDS data

Different RDS data were used at different points in our network simulation process. We estimated the joint distribution of characteristics of individuals in the PWID population to populate our initial network. We used RDS weights to calculate the distribution of important demographic and behavioral characteristics in the PWID population in each city.

The main characteristics of interest in this analysis were: current age; age of first injection drug use; injecting frequency; injecting in shooting galleries or tambayan (small, private gatherings among PWID); acquiring needles and syringes from the local city health office; and reported degree (the number of other PWID each person knows).

We also estimated homophily, a measure of assortative mixing on the network, to identify important mixing characteristics for our model. We explored the homophily parameters for several characteristics including age, age at first injection, injecting frequency, choice of location to inject (shooting gallery, "tambayan"), and getting needle and syringes through the city health office. We selected only those characteristics that showed significant homophily (i.e., values > 1.1 in at least one category). Population characteristics and homophily were calculated using the successive-sampling estimator [95], which was the default method in the RDS R-package, although estimates using the "Volz-Heckathorn RDS-II" estimator [97] did not differ from Gile estimator results, which was the default used in R.

Finally, clustering, as measured in terms of the RDS-weighted ego-network configuration distribution, was used to calibrate the ERGM coefficient parameters [99]. The ego-network was defined as a person (ego) and the people directly connected to him (alters). The ego-network configuration considered patterns of ties between alters. We treated each PWID as an ego with up to three alters: the person who recruited him and his two recruits. Our secondary questionnaire asked about ties between the person's recruiter and his recruits, but it did not directly observe ties between the two recruits, accounting for only 2 out of the 3 ties. In the absence of this information, we assumed that the probability of triad closure was the same rate as explicitly observed on the network, which was 0.87 in Cebu and 0.95 in

Mandaue. The configurations of individuals who recruited only one other PWID were treated as missing data and were excluded from the ego-network distributions.

Setting up the initial ERGM

Using the RDS-weighted joint distribution of characteristics and degree in each city, we generated random networks the size of the estimated PWID populations in each of our two cities: 3000 in Cebu and 1500 in Mandaue [120]. Each network also had a distribution of demographic and behavioral characteristics that mirrored the distributions found in our RDS-weighted samples from both cities. Using this random network and each node's reported degree, we randomly assigned ties between all nodes so that the degree distribution approximated a smoothed distribution around the observed in our data.

We then set initial values for coefficients on each parameter in the ERGM. Clustering parameters were initially set at reasonable arbitrary values, since they were later updated through subsequent model fits. Initial mixing parameters were estimated using a case-control approach, where cases were observed ties and controls were non-ties from an RDS-weighted sample. Both the ERGM and logistic regression models were used to estimate the odds of a tie, for different combinations of characteristics, but the ERGM generated those estimates conditional on the characteristics that defined the rest of the network. Thus, the ERGM could be configured to account for overall network properties, such as the degree distribution and triadic clustering.

The initial 'random' network we had constructed to this point had randomly assigned ties, but ensured that the degree distribution and demographic and behavioral characteristics were consistent with our RDS sample, and an ERGM formula with initial coefficient parameter values but had randomly assigned ties. We then simulated networks, using a

rewiring algorithm that rearranged a small number of ties between people on the network until the network realization was consistent with the initial ERGM coefficients. To reduce the sampling space and avoid model degeneracy, we constrained our network models to only those networks with the same degree distribution as the initial 'random' network.

Adjusting ERGM coefficients

The new networks simulated would have clustering and mixing that resemble the RDS-weighted data. However, our original mixing coefficients assumed a dyad-independent model – that is, each pair was independent of every other pair. The inclusion of a non-zero triadic clustering term in the ERGM indicated that the network was not dyad-independent. Therefore, we updated the homophily coefficients using logistic regression where 'cases' were observed ties from the RDS data and 'controls' were randomly sampled non-ties from the simulated network. With the specified model and updated homophily coefficients, we then simulated a new set of networks, starting from the last available network.

The ego-network configuration distributions were then compared to the empirical distribution, and the clustering parameter was tuned to more closely match the empirical ego-network configuration distribution. Changing the clustering parameter altered the homophily coefficients, so homophily was again updated before a new set of networks was simulated. This iterative process continued to minimize the difference between the ego-network distributions in the model networks and the empirical data, as measured by a chi-square statistic. For each city, we used the best fitting parameter estimates to simulate a minimum of 30 networks consistent with the ego-network configuration distribution in the observed data, based on the chi-square statistic. We then estimated summary parameters on each of these networks, including number of components and component size and average distance

between nodes. We also looked for nodes that were indicated as important in terms of one of three centrality measures: degree, betweeness, closeness [107-109].

Results

Recruitment and enrollment were conducted from May 25 to July 10, 2013 at the Cebu and Mandaue City Health Offices. Seven seed PWIDs were selected from each city, covering a range of demographic characteristics and geographies. Each city reached its respective target recruitment, reaching a total of 450 PWID respondents in Cebu and 303 in Mandaue. One seed from each city failed to recruit any friends into the survey. These two individuals were removed from the analysis. Two chains from each city recruited 70% of the survey respondents. Recruitment was slightly faster and more efficient in Cebu, while the survey in Mandaue had the longest recruitment chain, completing over 20 waves of recruitment.

Most demographic characteristics were similar across the two cities (Table 1). The average age of PWID was around 30 years in both cities, and most PWID did not attend or complete high school. Injecting behaviors in both cities were quite risky, though PWID in Cebu injected more frequently than PWID in Mandaue. A higher proportion of Cebu PWID visited shooting galleries, although both groups reported shooting galleries at greater than 75%. Cebu PWID also reported larger networks of PWID, though this could indicate a friendship network and may not directly correlate with the injecting network.

Strong homophily by age and injecting frequency was observed across both cities. In Cebu, ties within the youngest age group (15 to 24 years) occurred almost 50% more frequently than what would be expected at random (1.49). In both cities, ties within the oldest PWID (35 years and older) occurred almost one-third more frequently than expectation

(Table 5). In both cities, PWID who injected less than once a day had a strong preference for peers who injected at the same frequency. In Mandaue, PWID who injected more than three times per day had some heterophily or preference outside of their own group, but this may be an artifact of the small size of this group, which made up just 3% of the network. Finally, PWID in Cebu also had preferences for assortative mixing by utilization of needle exchange services at the city health office.

The simulated networks had the degree and homophily across most of the characteristics of interest consistent with the RDS data (**Error! Reference source not found.**). Characteristics and categories of age included in the model had homophily consistent with that observed in RDS data and were clearly different from the random network. Homophily by injecting group was more moderate than RDS estimates, but showed greater homophily than the initial random networks. Characteristics that showed no homophily in the RDS sample (e.g., first age of injecting drug use) also showed no homophily in our simulations, suggesting that the networks we had simulated were consistent with the networks that produced our RDS samples.

We did observe several notable differences between the network structure estimates in the two cities (Table 8). In Mandaue, simulation results suggested one large component connecting 65-80% of the estimated 1500 PWID, and hundreds of smaller, disconnected groups. In Cebu, over 98% of individuals on the simulated networks were connected to the largest connected component, and most simulations had fewer than 10 small, isolated components. The differences in component size suggested that the epidemic in Cebu had the potential to reach a larger proportion of the PWID population than in Mandaue, which was consistent with the higher HIV prevalence observed in Cebu. Furthermore, within the largest

components of each city, it was easy for PWID to reach each other, but PWID were more closely connected in Cebu, which had an average path length of 3.3 compared to 3.8 in Mandaue. The shorter paths in Cebu meant that infection can reach more people in a fewer number of steps, which could explain why we observed more rapid growth in HIV prevalence in Cebu than in Mandaue.

Discussion

In this paper, we simulated networks from RDS data to compare PWID network structures between Cebu and Mandaue, two cities in the Philippines, and to assess whether differences in those network structures could explain observed differences in the timing and spread of HIV infection in the two cities. Our results suggest that differences in connectivity among our simulated networks were consistent with the differences in HIV prevalence patterns observed in Cebu and Mandaue. In Cebu, PWID frequently visited shooting galleries to rent needles and inject drugs, which might explain why most individuals in this city's simulated networks were part of the largest connected component. In contrast, the largest connected component encompassed just 65-80% of the PWID population in Mandaue, with other PWID clustered in small groups of 10 or fewer PWID.

These findings are consistent with local observations, in which PWID in Cebu are known to frequent shooting galleries and those in Mandaue tend to inject in "tambayan" (private parties) with close friends they know. These small clusters prevent spread of infection and may explain why the introduction of HIV in Mandaue led to slower spread and resulted in lower HIV prevalence than in Cebu. Initially, PWID in Mandaue who were infected may have been part of these smaller isolated components, limiting the spread of HIV, and the observed rise in HIV prevalence might have occurred only after infection was

introduced into one or multiple PWID who were connected to the largest network component in this community.

Other network statistics in the two cities could explain why the PWID network would promote the more rapid spread of HIV infection in Cebu than in Mandaue. Average path length was decidedly shorter in Cebu than in Mandaue, despite the fact that Cebu has almost three times the number of PWID. One would expect that a larger network size would make it more difficult for any two people to reach each other. The unexpected shorter path length in Cebu may be a consequence of the differences in degree, as network density on the connected component in Cebu (0.007) was higher than in Mandaue (0.006). Mandaue also exhibited higher clustering than Cebu. Higher clustering creates redundant paths on the network, and may slow the spread of infection across the network [44].

Our simulations revealed several network differences that could explain the differences in observed prevalence patterns in the two cities, but we were unable to isolate individuals or subgroups who were particularly important to maintaining connectivity or short path lengths on each network. Such results may have been a consequence of the relatively high density of these networks. While reported degree was still within the range of PWID social networks [65,126], it exceeded the levels reported in drug sharing and injecting networks in recent studies [55,56,127].

A number of limitations should be considered when assessing our results and conclusions. First, while the networks here closely matched the degree distributions and homophily observed on our true network, the ego-networks collected data on only 3 alters and only 4 possible configurations. These data were still used to estimate the clustering parameters in our model. Future studies that employ respondent driven sampling could

incorporate questions about each respondent's ego-network to describe networks and consider their potential impacts on the spread of infection. Second, our data were based on social networks and assumed that PWID share drugs or needles with all of their social ties. While this may have overstated the true state of drug and needle-sharing, the drug and injection networks would be a subset of those simulated. Additional knowledge of the direct risk network may offer a better estimate of how rapidly and broadly infection would spread in this community.

PWID remain a vulnerable population in the global HIV epidemic. After 30 years of the global HIV epidemic, new outbreaks of HIV among PWID continue to emerge throughout the world [1-4,128,129]. In the work presented here, we made use of routine surveillance data to identify differences in network connectivity that may explain observed differences in disease dynamics in PWID populations. We identified that differences in the size of the largest connected component and average path lengths could result in differences in timing and speed of HIV spread. Future work would consider fitting a transmission model to evaluate how introducing infection along different parts of the network might have impacted the speed and extent of infection in the population, and how to identify individuals on the network who may be important for preventing widespread infection. These models might also be used to simulate the impacts of cutting off or slowing spread of infection through prevention or treatment programs. A better understanding of the interplay between network structures and disease dynamics could play an important role in identifying effective intervention and thus strengthen our progress towards reducing the spread of HIV.

Figures and Tables

Table 4. Demographic and behavioral characteristics of the sample, weighted using Gile's successive sampling estimator [95].

	Cebu		Mandaue	
	%	95% CL	%	95% CL
Age group				
15 to 24	24	(22 - 34)	42	(29 - 54)
25 to 34	32	(25 - 39)	27	(18 - 36)
35 and older	40	(33 - 47)	32	(23 - 41)
Frequency of Injecting				
< 1x per day	47	(39 - 54)	64	(53 - 74)
1-2 x per day	20	(15 - 25)	28	(17 - 38)
2-3 x per day	19	(14 - 24)	5	(2 - 9)
>3 x per day	14	(10 - 19)	3	(1 - 6)
Age at first injection drug use				
Under 18	34	(27 - 41)	36	(24 - 49)
18 to 25	40	(33 - 47)	46	(35 - 58)
Over 25	26	(19 - 33)	2	(9 - 25)
Places to inject[1]				
Shooting galleries	84	(78 - 89)	84	(78 - 90)
Tambayan	24	(17 - 30)	20	(13 - 26)
Sources for needles[2]				
City Health Office	16	(12 - 21)	4	(0 - 8)
Pharmacy	24	(17 - 30)	24	(13 - 34)
Other PWID	16	(11 - 22)	27	(17 - 38)
Shooting gallery	43	(36 - 50)	44	(33 - 56)

1 Participants were asked about injecting at each location separately, so percentages to not sum to 100%

2 Participants were asked to name all places they accessed needles, so percentages may not sum to 100%

Table 5. Homophily of characteristics of interest.

	Cebu	Mandaue
Age group		
15 to 24	1.487	1.177
25 to 34	1.167	1.007
35 and older	1.321	1.304
Frequency of Injecting		
< 1x per day	1.262	1.157
1-2 x per day	1.083	0.983
2-3 x per day	0.980	1.380
>3 x per day	1.072	0.823
Age at first injection drug use		
Under 18	0.990	0.967
18 to 25	0.996	0.958
Over 25	1.142	0.977
Places to inject[3]		
Shooting galleries	1.077	1.124
Tambayan	0.999	1.259
Sources for needles[4]		
City Health Office	1.372	1.012
Pharmacy	1.025	1.061
Other PWID	0.884	1.121
Shooting gallery	1.124	0.968

3 Participants were asked about injecting at each location separately

4 Participants were asked to name all places they accessed needles

Table 6 (a/b). Comparison of RDS-weighted characteristics and random network characteristics in Cebu (a) and Mandaue (b).

Cebu	Proportions		Degree		
	RDS-weighted % (95% CL)	Random network %	RDS-weighted Mean (95% CL)	Random network Mean	Final network solution Mean (95% CL)
Age group					
15 to 24	24 (22 - 34)	28	20.0 (15-25)	20.0	21.58 (21.08-22.07)
25 to 34	32 (25 - 39)	32	21.0 (15-27)**	20.7	20.68 (20.23-21.12)
35 and older	40 (33 - 47)	40	21.4 (16-27)	21.5	21.76 (21.33-22.18)
Frequency of Injecting					
< 1x per day	47 (39 - 54)	47	20 (18-23) **	20.2	20.55 (20.30-20.79)
1-2 x per day	20 (15 - 25)	20	21 (15-28)	21.3	21.92 (21.10-22.73)
2-3 x per day	19 (14 - 24)	19	21 (15-28)	21.1	23.09 (22.25-23.93)
>3 x per day	14 (10 - 19)	14	22 (13-30)	21.8	20.96 (20.27-21.66)
Age at first injection drug use					
Under 18	34 (27 - 41)	34	22 (16-27)	21.4	22.25 (22.00-22.49)
18 to 25	40 (33 - 47)	40	23 (19-28)	23.2	22.76 (22.40-23.11)
Over 25	26 (19 - 33)	26	17 (11-22)	16.4	18.03 (17.52-18.55)
Places to inject					
Shooting galleries	84 (78 - 89)	84	21 (18-25)	21.4	21.71 (21.64-21.78)
Tambayan	24 (17 - 30)	24	21 (18-25)	19.4	20.98 (20.74-21.23)
Sources for needles					
City Health Office	16 (12 - 21)	16	25 (17-24)	24.4	30.23 (29.33-31.12)
Pharmacy	24 (17 - 30)	24	23 (16-29)	22.9	20.92 (20.52-21.32)
Other PWID	16 (11 - 22)	16	19 (14-24)**	18.8	18.74 (18.21-19.26)
Shooting gallery	43 (36 - 50)	43	22 (18-26)	19.1	19.34 (18.94-19.74)

Mandaue	Proportions		Degree		
	RDS-weighted % (95% CL)	Random network %	RDS-weighted Degree (95% CL)	Random network	Final network solution
Age group					
15 to 24	42 (29 - 54)	42	7 (4-9)	6.92	6.77 (6.63-6.90)
25 to 34	27 (18 - 36)	27	10 (7-13)	9.65	10.33 (10.02-10.64)
35 and older	32 (23 - 41)	32	9 (6 -13)	9.21	8.84 (8.67-9.01)
Frequency of Injecting					
< 1x per day	64 (53 - 74)	64	7 (5-10)	7.46	7.35 (7.27-7.43)
1-2 x per day	28 (17 - 38)	28	9 (6-12)	8.85	9.22 (9.14-9.30)
>2 x per day	9 (5 - 13)	9	14 (6-22)	13.6	13.22 (12.61-13.83)
Age at first injection drug use					
Under 18	36 (24 - 49)	36	8 (6-11)	8.30	8 (7.94-8.40)
18 to 25	46 (35 - 58)	46	9 (6-12)	9.13	9 (8.75-9.08)
Over 25	2 (9 - 25)	17	7 (4-9)	6.54	7 (7.09-7.66)
Places to inject					
Shooting galleries	84 (78 - 90)	84	8 (6-10)	7.75	7.50 (7.43-7.56)
Tambayan	20 (13 - 26)	20	13 (8-18)	12.7	13.50 (13.23-13.76)
Sources for needles					
City Health Office	4 (0 - 8)	4	9 (5-12)**	8.7	7.30 (6.69-7.91)
Pharmacy	24 (13 - 34)	24	9 (5-12)	8.8	8.81 (8.64-8.98)
Other PWID	27 (17 - 38)	27	9 (5-12)	8.5	9.28 (9.03-9.52)
Shooting gallery	44 (33 - 56)	44	8 (5-11)	8.7	7.60 (7.47-7.73)

** Confidence limits could not be estimated using RDS methods; numbers presented use weighted survey methods and are narrower than RDS estimates

Table 7. Comparison of homophily from RDS, random network and the final solution shows homophily from our solutions are somewhat close to empirical estimates.

Homophily	Cebu RDS	Cebu Random	Cebu Final solution mean (range)	Mandaue RDS	Mandaue Random	Mandaue Final solution
Age group						
15 to 24	1.487	1.012	1.19 (1.15-1.24)	1.177	1.022	1.15 (1.12-1.18)
25 to 34	1.167	1.008	1.13 (1.12-1.15)	1.007	0.992	1.02 (1.01-1.04)
35 and older	1.321	1.004	1.25 (1.23-1.28)	1.304	1.012	1.29 (1.25-1.35)
Frequency of Injecting††						
< 1x per day	1.262	0.996	1.04 (1.03-1.05)	1.157	1.016	1.07 (1.05-1.09)
1-2 x per day	1.083	1.002	1.10 (1.09-1.12)	0.983	1.015	1.01 (0.99-1.03)
2-3 x per day	0.980	0.999	1.00 (0.99-1.01)	1.167	1.006	1.08 (1.06-1.09)
>3 x per day	1.072	0.993	1.03 (1.02-1.05)	--	--	--
Age at first injection drug use						
Under 18	0.990	1.000	1.03 (1.02-1.04)	0.967	1.014	1.01 (1.01-1.03)
18 to 25	0.996	1.003	1.01 (1.00-1.02)	0.958	1.005	1.00 (0.99-1.01)
Over 25	1.142	0.995	1.04 (1.03-1.05)	0.977	1.000	1.03 (1.02-1.05)
Places to inject						
Shooting galleries	1.077	0.997	1.00 (1.00-1.01)	1.124	0.988	1.08 (1.06-1.10)
Tambayan	0.999	0.996	1.01 (0.99-1.02)	1.259	0.986	1.13 (1.11-1.15)
Sources for needles						
City Health Office	1.372	1.007	1.04 (1.03-1.07)	1.012	1.005	1.04 (1.02-1.05)
Pharmacy	1.025	0.993	1.00 (0.99-1.01)	1.061	0.994	1.02 (1.00-1.04)
Other PWID	0.884	0.993	1.01 (1.00-1.01)	1.121	1.004	1.04 (1.03-1.05)
Shooting gallery	1.124	0.995	1.01 (0.99-1.02)	0.968	1.001	1.04 (1.02-1.06)

†† The higher two categories of injection frequency were collapsed in Mandaue network estimates and simulations due to small cell sizes

Table 8. Summary network statistics of simulations in Cebu and Mandaue.

Network Statistics	Mandaue Mean (range)	Cebu Mean (range)
Number of components	132 (127-138)	6 (2-13)
Largest component (size)	1082 (1069-1100)	2981 (2955-2995)
Density	0.00559	0.00712
Average path length	3.774 (3.689-3.972)	3.302 (3.268-3.349)
No. triangles	13,480 (13,297-13,596)	129,814 (127,214-133,438)
Clustering Coefficient	0.386 (0.381-0.389)	0.293 (0.287-0.301)
Centrality (Individual statistics)		
Degree	10.8 (10.7-10.9)	21.5 (21.4-21.6)
Eigenvector	0.020 (0.019-0.020)	0.011 (0.011-0.012)
Betweenness (information)	2998.1900 (2872-3257)	78.2711 (3257-0)
Closeness	0.270 (0.260-0.276)	0.308 (0.304-0.311)
Centralization (network statistics)		
Degree	0.069 (0.068, 0.070)	0.085 (0.084, 0.085)
Eigenvector	0.177 (0.165, 0.203)	0.158 (0.143, 0.172)
Betweenness	0.075 (0.068, 0.091)	0.055 (0.048, 0.062)
Closeness	0.110 (0.102-0.115)	0.132 (0.123-0.138)

CHAPTER 5: A PHYLOGENETIC ANALYSIS OF EMERGING EPIDEMICS AMONG PEOPLE WHO INJECT DRUGS IN THE PHILIPPINES

Introduction

People who inject drugs (PWID) remain central to the global HIV epidemic. Although they comprise just 0.26% of the world's population [14], they disproportionately account for one-third of new HIV infections outside of sub-Saharan Africa [130]. HIV epidemics among PWID can reach very high proportions, in some cases infecting over one-third of the population at risk [12,128,131,132]. Furthermore, HIV epidemics among PWID expand rapidly. HIV prevalence may increase from virtually 0% to 50% within less than one year [6], and new epidemics continue to emerge among PWID [1,128,129]. Behavioral surveys can be used to identify settings where outbreaks may occur, but they cannot predict the timing and onset of an epidemic. This unpredictable nature and the potentially severe consequences of these HIV epidemics, PWID should be a top priority for HIV prevention and care.

In this work, we used phylogenetic analysis to investigate the emergence of HIV among PWID located in two cities in the Philippines. In Cebu City, active surveillance documented a rapid rise in HIV prevalence among PWID -- from less than 1% in 2009 to 53% in 2010 [11,12]. This was followed two years later by a similar rise in HIV prevalence among PWID in the neighboring city of Mandaue, where prevalence grew from 3.6% in 2011 to 38% in 2013 [11-13]. Phylogenetic trees of HIV sequences, collected through surveillance in the initial stages of the epidemic in Cebu City showed that 90% of infections among PWID arose from a common source [133,134]; however, during the same period, only 11

HIV infections were detected among injecting drug users in Mandaue [12], and their sequences could not be amplified for further phylogenetic analysis.

The observed emergence of HIV in these two cities allows us to examine potential injection sharing links between PWID in Mandaue and Cebu. If PWID in Mandaue were well-integrated with the Cebu injecting network, we would not expect a 2-3 year delay before the emergence of HIV in Mandaue. One possible explanation for the delay is that PWID in Mandaue are still connected Cebu injecting network, but they are connected to the periphery, so that infection is slower to reach them. Alternatively, the needle-sharing networks in Cebu and Mandaue are completely separated, and the Mandaue epidemic was a result of an external introduction of infection into Mandaue's high-risk injecting network.

To determine whether the emergence of HIV among PWID in Mandaue was connected to PWID in Cebu, we assessed the genetic clustering of HIV infections collected in active HIV surveillance in these two cities. We constructed a phylogenetic tree of the genetic sequences collected from PWID who participated in the 2013 HIV surveillance surveys in Cebu and Mandaue. We then used this tree to identify clusters of infection and to determine whether the Mandaue epidemic arose from the same source as the epidemic in Cebu.

Methods

Sampling and enrollment

The data analyzed were a subset of the PWID surveys collected by the Integrated HIV Behavioral and Serological Surveillance (IHBSS) program in 2013. IHBSS surveys have been conducted every two years by the Philippines Department of Health to assess the HIV risk in the country for HIV strategic planning. They have surveyed several high-risk

populations in a number of cities across the country. Behavioral questionnaires were administered via face-to-face interviews to measure risk behaviors, and blood was drawn to evaluate for Hepatitis C (HCV), syphilis and HIV. All data were collected anonymously and surveys were not subject to institutional review as the work was part of ongoing surveillance. The analyses presented here were deemed exempt by the UNC IRB.

In the 2013 IHBSS, recruitment of PWID was conducted using respondent-driven sampling (RDS), a strategy often employed to identify and recruit individuals when the sampling frame is not well defined. RDS is particularly useful for recruiting stigmatized populations, because participating members of the community are asked to recruit their peers directly, which may be more persuasive and protects the confidentiality and anonymity of those who refuse to participate. The RDS theory and method have been described previously [93,135]. In our surveys, an initial group of PWID (referred to as "seeds" in RDS terminology) were identified and asked to recruit 2-3 friends whom they knew to inject drugs. Successfully recruited PWID completed the survey and were subsequently asked to recruit 2-3 more of their friends who also inject drugs. This process was repeated with each subsequent round of recruits until the target sample sizes of 450 and 300 recruits were reached in Cebu and Mandaue, respectively.

Next-generation HIV sequencing

Blood samples were tested for HIV at the STD/AIDS Surveillance Cooperative Central Laboratory (SACCL), San Lazaro Hospital, Manila. Plasma RNA was extracted from all HIV-positive samples, using the QiAmp Viral RNA Isolation Kit (Qiagen, Germany) and shipped to the Hawaii Center for AIDS Research (HFCA) in Honolulu, Hawaii (USA). To minimize RNA degradation during transport, the samples were stabilized into lyophilized

form using GenTegra RNA Matrix (GenTegra, USA). Upon arrival at HCFA, the RNA was reconstituted in RNAse-free water according to manufacturer protocol and used as a template for amplicon generation.

As the primary purpose of the HIV sequencing was to inform Ministry of Health treatment guidelines, the main target for replication was part of the p51 (Reverse Transcriptase) region of the *pol* gene (HXB2 2748-3216). Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) was conducted using a one-step RT PCR protocol, which synthesizes first-strand cDNA and subsequent PCR in a single reaction, using MyTaq™ One-Step RT-PCR Kit (Bioline, USA). Fusion primers that were designed to match the 454 Roche sequencing system were gene-specific sequences appended with 454-template adapter sequences and a short sequence tag (or Multiple Identifier) to uniquely identify each amplicon.

A 1.5 μL solution of the PCR reaction was run on a 48-lane E-Gel Agarose Gel Electrophoresis system (Life Technologies, USA) to visually inspect for the HIV *pol* amplicon band, which indicated amplification success. In samples where amplification was successful, the remaining PCR product was purified using Agencourt AMPure XP beads (Beckman Coulter, USA) and quantified using Quant-IT PicoGreen dsAssay kit (Life Technologies, USA), followed by subsequent normalization and pooling.

Next-generation sequencing (NGS) was implemented using the GS Junior sequencing system (Roche 454 Life Sciences, USA). Emulsion-based clonal amplification PCR (emPCR) was performed to clonally replicate each sequence recovered in the RT-PCR process. Amplicons were individually sequenced from forward and reverse ends, to generate up to 200,000 reads per sample. Data cleaning was completed using Intergroomer

(http://courge.ics.hawaii.edu/inte/groomer/) and consensus sequences were determined from each sample, based on the majority nucleotide base at each site.

Phylogenetic Analysis

Phylogenetic trees were constructed using the consensus sequences from HIV samples in the 2013 IHBSS survey and reference sequences from three sources. The first two sets of reference sequences were taken from GenBank. We pulled any sequences in GenBank tagged as originating from the Philippines. In addition, we also retrieved the ten sequences most genetically similar to our 2013 sequences of interest, which was found using the HIV BLAST search tool (http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). The third set of reference sequences consisted of all successfully amplified HIV sequences acquired from 2013 IHBSS surveys of other cities and populations in the Philippines. These reference sequences were included to calibrate the degree of clustering in the sequences we observed. If our sequences of interest clustered with reference sequences, then we could not conclude that the HIV infections detected in Cebu and Mandaue in 2013 were any more closely related to each other than to other "random" sequences. In other words, such a result would suggest that the data did not exhibit any genetic similarity within our sequences of interest, beyond the similarity expected by random chance.

All sequences were aligned to the HXB2-LAV K03455 reference genome using MUSCLE [112,113]. Using this alignment file, a maximum likelihood (ML) tree was constructed in RAxML [114] using a generalized time-reversible (GTR) gamma model of substitution with rate heterogeneity [58]. We conducted runs with 1,000 bootstrap samples. Transmission clusters were identified as clades with short branch length and high bootstrap

support (95%) [115]. We described the size and composition of transmission clusters that included PWID sequences from the 2013 IHBSS in Cebu or Mandaue, and assessed whether there was any apparent separation between sequences from Cebu and those from Mandaue.

The consensus sequences from the 2013 were deposited in GenBank with accession numbers KX184117-KX184197. Other sequences used in the phylogenetic tree were accession numbers AB587097-AB587118, AB747376-AB747549, KF059718-KF059833, KJ868976-KJ868980.

Statistical Analysis

Characteristics of the HIV-infected populations were estimated for each city using RDS sequential-sampling weights [95]. Weights were first constructed for each city, using local population size estimates of 4500 PWID total, and subset by each city: 3000 in Cebu and 1500 in Mandaue. Demographic and behavioral characteristics were reported for only the subset of HIV-infected PWID, and stratified by city and amplification status. Statistical tests for differences were conducted using the Rao-Scott test for survey-weighted data [136]. All analyses were conducted using the 'RDS' package in R Software [121,122]. There is little consensus on the selection and use of RDS-weights for estimation [110,137-139]; we report weighted results for this work, but note that unweighted results did not show appreciable differences.

Results

Study population

The IHBSS survey achieved the target sample sizes in both surveys, recruiting a total of 457 male PWID in Cebu and 310 male PWID in Mandaue, including the initial wave of seed PWID. We restricted our analyses to those PWID who tested HIV-positive, so that our

study population comprised 239 male PWID from Cebu and 131 from Mandaue. Surveillance staff identified and excluded 33 HIV-infected participants in Mandaue who had already participated in Cebu, using a unique identifier (using both parents' initials, birth order, and date of birth) collected on the enrollment form. After removing duplicate samples, PCR amplification and sequencing was successful in 92 (27%) unique male PWID samples: 68 from Cebu and 24 from Mandaue.

Demographic and behavioral characteristics were similar in the amplified and non-amplified populations and in the two cities (Table 9). Injectors in both cities were on average around 30 years of age, and had been injecting for 11 years. Most had completed some schooling, but only one-third had completed at least a high school education. All groups reported equally risky injecting practices. A majority of injectors reported sharing needles in the last 6 months, and over 60% reported injecting with a used needle. While injectors in Cebu reported more shooting gallery visits than those in Mandaue (7 vs 3), almost 80% in both cities reported injecting at a shooting gallery most of the time and over 80% had pooled funds with other drug users to purchase drugs.

Phylogenetic Tree Analysis

The BLAST search of geographically and genetically similar sequences generated a total of 286 unique reference sequences. These were combined with 92 sequences from male PWID in Cebu and Mandaue (68 from Cebu, 24 from Mandaue) and another 17 sequences from IHBSS surveys among other risk groups (6 from female PWID and 11 from MSM). A total of 395 sequences were included in the phylogenetic analysis.

The resulting phylogenetic tree showed very high bootstrap support (97%) for a large subtype B cluster of 172 sequences, which included 99% (N= 92) of infections among the

male PWID infections and all 6 of the female PWID infections in Cebu and Mandaue in the 2013 surveys (Figure 7). One sequence sampled from the 2013 MSM survey was also in this cluster, but the remaining 10 were scattered throughout the tree. The cluster was predominantly composed of sequences obtained from PWID (Table 10). The 2013 PWID sequences made up over half (52%) of the sequences in this cluster; and an additional 49 sequences (28%) in the cluster were from prior surveillance surveys of PWID in Cebu or case reports from people reporting injecting drug use (shown in purple in Figure 7). Two sequences each were from prisoners and MSM; and three sequences were sampled from female sex workers (FSW). We were unable to verify the risk factors related to nineteen of the sequences from the cluster.

Only one sequence from a male PWID in the 2013 survey fell outside the large cluster. That sequence clustered with CRF-01 AE subtype, which has been reported in a number of MSM infections in the Philippines [134]. The CRF01-AE infected PWID was older than average (40 years) and started injecting drug use at age 35.

Discussion

We have presented a phylogenetic analysis of the emergence of HIV in two distinct patterns across neighboring cities in the Philippines. We found high bootstrap support for a large subtype B cluster of predominantly PWID infections (84%), including almost all HIV infections among PWID in the recent 2013 IHBSS round. The high bootstrap support of a monophyletic clade of HIV sequences that included almost all our PWID samples suggests that new infections in Mandaue arose from overlapping needle-sharing networks of PWID in

the two cities. In contrast, MSM infections showed no genetic relatedness, as sequences were widely scattered throughout the tree, consistent with prior analyses [134].

We found low bootstrap support for branching patterns within the large cluster, which may indicate that sequences lacked the genetic diversity needed to draw any distinctions between them. The lack of a clear separation by city suggested that the new infections emerging in Mandaue were not a result of a single introduction, leading to spread across a risky injection-sharing network in Mandaue. Instead, it was more likely a result of repeated contact and needle sharing between PWID in Cebu and Mandaue. Such repeated exposures may be through direct contact or sharing with PWID in Cebu or through shooting galleries, which were associated with higher HIV infection [140,141]. In Mandaue, almost three-quarters of PWID surveyed inject at a shooting gallery, and of those, 73% attended shooting galleries in Cebu.

While we found an interesting connection between epidemics in Cebu and Mandaue, several limitations should be considered in interpreting these results. A large number of infections could not be successfully amplified. Amplification was performed on plasma RNA, which is highly sensitive to degradation by RNAse. The rapid temperature changes, combined with possible contamination from blood processing done in the field may have impacted the degree of degradation by the time sequencing was conducted. Poor amplification results in missing data, which may bias our results. In the extreme, if amplification succeeded only for PWID infections that fall within the cluster of interest, it may be possible that the data we were unable to collect represent a completely separate transmission cluster that we did not observe. Comparisons of measured demographic and

behavioral characteristics were quite similar across the two groups, supporting the hypothesis that missing sequences were likely similar to those that we observed.

Moreover, even if the sequences were representative of our sample, the length of the sequences analyzed may have been insufficient to observe important and relevant mutations. Our reads were about 400 base pairs in length, which might not provide sufficient detail to discern sub-clusters. One way to extract greater detail would be to sequence a longer portion of the viral genome. However, given the very short time that elapsed between the onsets of HIV spread in these two groups, even whole genome sequencing could be insufficient to distinguish between differences that occur within the host versus those that occur across different hosts.

Finally, our tree clusters could have be confounded by antiretroviral therapy (ART), which exerts selective pressures that could induce greater genetic distance between linked infections. Although the survey did not directly collect data on the treatment status of its participants, it was unlikely that this would bias our analysis as fewer than 30 PWID were on ART in the whole province at the time of surveillance in 2013 [National Philippine Surveillance System].

Conclusions

Using phylogenetic methods to describe the clustering patterns of HIV from two epidemics in the Philippines, we have shown that the two epidemics arose from a common source of infection, which suggested substantial needle sharing between drug users in both cities. Understanding the transmission dynamics of HIV infection among people injecting drugs is particularly important in the context of emerging epidemics. In light of the clustering

we observed, one should consider PWID in Cebu and Mandaue as a single connected population, with shared contact and exposure, instead of approaching each separately. This approach to the population may guide strategies for providing successful HIV prevention to people injecting drugs in both cities.

Figures and Tables

Figure 7. Best maximum likelihood (ML) phylogenetic tree of all 395 sequences in the Philippines with 1,000 bootstrap replicates. Branches are colored by mode of transmission. The large cluster of 172 sequences has a bootstrap support of 97% and contains almost all sequences from PWID in Cebu and Mandaue.

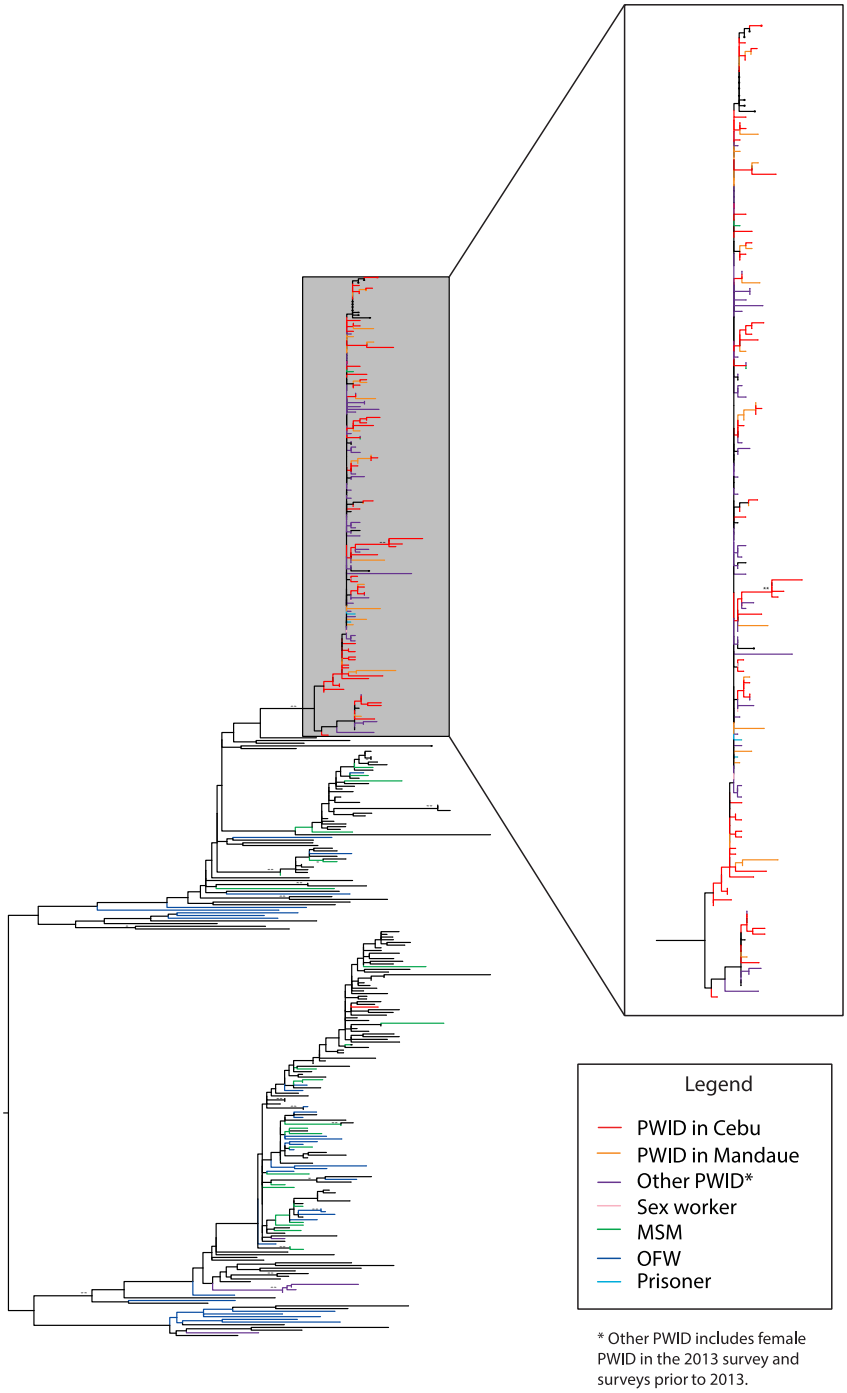

Table 9. Comparison of behavioral characteristics, by PCR amplification status, in Cebu and Mandaue.

	Cebu					Mandaue					Total				
	PCR-		PCR+			PCR-		PCR+			PCR-		PCR+		
	N‡‡	(%)	#	(%)	p	#	(%)	#	(%)	p	#	(%)	#	(%)	p
Sample Size (N)	172		67			102		29			*274*		*96*		
Weighted Sample§§	1166		385			351		167			*1300*		*536*		
Mean age at enrollment (sd)	33 (0.9)		33 (1.5)		0.97	34 (1.3)		31 (1.8)		0.18	34 (0.8)		32 (1.3)		0.23
Education					0.09					0.12					0.08
No grade completed	275	(24)	59	(15)		60	(17)	6	(4)		263	(20)	46	(9)	
Elementary	377	(33)	202	(52)		211	(60)	105	(63)		603	(46)	315	(59)	
High School	484	(42)	104	(27)		71	(20)	50	30		401	(31)	154	(29)	
Vocational/College	24	(2)	20	(5)		9	(2)	5	(3)		31	(2)	21	(4)	
Injection Sharing Practices					0.07					0.09					0.02
Shared last injection	315	(27)	87	(23)		129	(37)	25	(15)		409	(31)	97	(18)	
Shared but not last injection	560	(48)	121	(31)		80	(23)	30	(18)		469	(36)	126	(24)	
Never in the past 6 months	291	(25)	177	(46)		142	(41)	111	(67)		422	(32)	313	(58)	
Where do you usually inject drugs?					0.51					0.61					0.85
At a shooting gallery	1046	(90)	322	(84)		269	(77)	131	(79)		1087	(84)	434	(81)	
In a "tambayan"	119	(10)	57	(15)		73	(21)	36	(21)		197	(15)	99	(18)	
Other	0	(0)	6	(2)		9	(2)	0	(0)		16	(1)	4	(1)	
Additional Injecting Practices															
Used Needle First	426	(37)	135	(35)	0.86	148	(42)	37	(22)	0.09	508	(39)	147	(27)	0.12
Injected with a used needle	793	(68)	263	(68)	0.99	202	(58)	132	(79)	0.06	820	(63)	399	(75)	0.12
Pool funds to buy drugs	1067	(92)	303	(79)	0.05	282	(81)	149	(89)	0.31	1119	(87)	454	(85)	0.77
Age at injection drug use (sd)	23 (0.8)		21 (1.1)		0.33	21 (1.0)		22 (1.5)		0.59	22 (0.8)		22 (0.6)		0.95
Frequency of drug use															
Injections in the last day	3 (0.2)		3 (0.3)		0.52	3 (0.2)		3 (0.4)		0.43	3 (0.2)		3 (0.1)		0.63
Injections in the last week	15 (1.3)		14 (2.0)		0.69	9 (0.9)		8 (2.1)		0.69	9 (0.9)		12 (0.9)		0.36
Shooting gallery visits	8 (0.7)		7 (1.1)		0.28	3 (0.4)		3 (0.8)		0.92	3 (0.4)		6 (0.5)		0.16

‡‡ All numbers are weighted sample sizes

§§ All weights were constructed using sequential-sequencing RDS weights

Table 10. Composition of the main transmission cluster (N=172), by HIV exposure category.

Exposure category	*2013 IHBSS*	*Reference*	*Total*
People who Inject Drugs (PWID)	97	49	146
Male	*91*	-	-
Female	*6*	-	-
Men who have Sex with Men (MSM)	1	1	2
Sex Worker (SW)	0	3	3
Prisoner	0	2	2
Unknown	0	19	19
Total	98	74	**172**

CHAPTER 6: DISCUSSION

PWID make up less than 1% of the global population [14], yet they comprise almost one-third of new HIV infections outside of sub-Saharan Africa [130]. While spread of infection in heterosexual epidemics has slowed in most of the world, new HIV epidemics among PWID continue to emerge [2-4]. To address prevention for PWID populations, we have an extensive understanding of behavior change approaches to limit risk of infection given exposure to the virus; but we still know little about how to assess one's risk of exposure – and more specifically, how network structures could determine one's risk of have an HIV-infected injecting partner. In this dissertation, we consider network structures among PWID and their potential impacts on disease dynamics.

This dissertation examined patterns in the emergence of HIV among PWID in two cities in the Philippines. In Cebu City, we observed a rapid rise in HIV prevalence, from 0.6% in 2009 to over 50% in 2010 [11,12]. In Mandaue City, the growth in HIV was delayed and grew at a slower pace, with HIV prevalence initially documented at 3.5% in 2011 and growing to 38% in 2013 [11-13]. We used exponential random graph models to simulate the underlying social networks in each of the two cities and consider whether differences in network structures could explain the differences in the speed and extent of disease transmission in the two cities (Aim 1). We also generated a phylogenetic tree of all available HIV infections in the Philippines to assess the extent to which HIV infections among PWID in the two cities were linked (Aim 2).

Aim 1: Evaluating the potential impact of network structures among PWID on HIV transmission

Summary of Findings

In Aim 1, we simulated networks that had homophily and clustering characteristics that were consistent with what we observed in our RDS-weighted data. Several characteristics of the resulting networks could explain some of the differences in disease dynamics that we observed. First, networks suggested that PWID in Cebu were almost completely connected, with the largest connected component including 2981 (range: 2955, 2995) of the 3000 PWID in this city. In network simulations for Mandaue, the largest connected component included only 1082 (range: 1069, 1100) of 1500 PWID. If spread of infection was limited to ties in our networks, then the largest connected components would delineate the reach of disease spread and the differences we observed could have explained why HIV prevalence reached higher levels in Cebu than Mandaue. This conclusion was consistent with other network studies, which found that the presence of HIV infection on disconnected parts of the network could explain the relatively low HIV prevalence in otherwise high-risk networks [64,142].

Several other network parameters differed across the two cities. Average degree was considerably higher in the high HIV prevalence Cebu than Mandaue (21.5 v 10.8). These results were consistent with transmission model simulations, which found that the probability of an outbreak and rate of its spread would rise with increasing degree [28]. In addition, higher clustering was observed in Mandaue than in Cebu. At all levels, clustering increased the number of redundant pathways between to individuals by adding a direct connection between two people who were already connected through one or more mutual intermediate friends. At low levels of clustering, an additional redundant path could increase cohesion on

the network, promoting spread of infection by making the network less vulnerable to random individuals dropping out of the network or other random breaks in transmission. At higher levels, like those observed in Cebu and Mandaue, an increase in redundant pathways could slow disease transmission, even confining transmission into isolated and highly clustered sub-networks [39,40,44]. The combined effects of these differences in network characteristics were consistent with the observed faster spread of HIV in Cebu than in Mandaue.

Public Health Significance

Network structures may play an important role in disease transmission by creating separation between components that could prevent spread or by increased ties that could create more opportunities for acquisition and onward transmission. In this work, we used HIV surveillance data, which have been routinely collected, to simulate and estimate network properties among PWID in two cities. We found differences in four major network characteristics that may explain the variation in HIV prevalence patterns in the two cities. First, the sizes of the largest connected component were important to defining the reach of disease spread. Cebu had a much larger component that also connected over 95% of network members and therefore had a higher potential for overall disease transmission. Second, average path length was shorter in Cebu than Mandaue, which suggests that infection in Cebu could reach a larger number of individuals on the network in a fewer number of steps. Third, higher clustering in Mandaue could indicate the creation of highly clustered sub-networks that provide areas that may constrain the spread of infection. Finally, higher average degree in Cebu could explain the more rapid spread and higher prevalence level by shortening average path length and broader higher connectivity on the network.

The most striking difference between the two cities was the number and size of connected components. The numerous small isolated components in Mandaue may have slowed the introduction of infection into the largest connected component, which could explain the delay in increased HIV infection there. Such findings may have limited implications for prevention. It would be challenging to design interventions that could or limit contain injection sharing practices to small groups to mimic the small, isolated network components that we observed in Mandaue. However, these findings did expand our understanding connectivity among these PWID populations. These methods could be used to identify and prioritize those populations with connectivity patterns similar to what we observed in Cebu, which would have the greatest potential for rapidly expanding outbreaks and the most urgent need for prevention efforts.

While the methods used in this work were computationally intensive which may at present limit its applicability in non-research settings, our findings may suggest that several useful network properties could be easily measured directly in data without the need for complex computer simulations. For example, average degree is collected as a part of any respondent-driven sampling survey. Local clustering around an individual could be approximated through questions about his friends and their connections with each other [99]. If network-level characteristics can be identified through data collected on a sample of individuals, we may be able to use surveillance data to quickly assess and understand the epidemic potential of a community.

Limitations

The network reconstruction simulations had several limitations. Characteristics of people on the network were calculated using RDS weighted network estimates, which have

been often controversial and may not be unbiased [143]. The empirical ego-networks were incomplete, so as we fit the clustering parameters, we compared and adjusted the results according to a truncated distribution, which is incomplete and may allow for a wide range of sufficiently fitting models. Moreover, network simulations were based on reports of social ties, and thus do not necessarily define injection risk on the network. However, these social ties probably represent the broadest or widest network, and the qualitative differences in degree, path length and component size across the two cities would hold for the injecting network as well.

Future Research

A natural next step to this work would be to apply an agent-based transmission model to the network. If the transmission model could substantiate the observed differences in HIV prevalence in the two cities, we would have greater confidence in our constructed networks. Such a model would also allow us to directly assess how network structures or sub-structures impact disease transmission, by comparing how disease dynamics change as we alter the network environment surrounding the initial infected person. For example, we could demonstrate how multiple introductions of infection into the disconnected parts of the network would limit the spread of the epidemic, and could explain the slowed spread of infection in Mandaue. Our models may further investigate how the transmission dynamics might change if the initial infection was introduced on highly clustered network sub-structures, or if targeted prevention programs were able to stop transmission among centrally located individuals on the network. Overall, refinement and improvement of the network simulation and epidemic models would provide a more robust tool for further investigation of the effects of changes in network structure on epidemic spread.

Aim 2: A phylogenetic analysis of emerging epidemics among people who inject drugs in the Philippines

Summary of Findings

In Aim 2, we conducted a phylogenetic analysis of the HIV infections detected in the two cities to assess whether the emergence of HIV in Mandaue was linked to the ongoing epidemic in Cebu. We constructed a phylogenetic tree using all the 2013 surveillance sequences, combined with a number of references sequences. The resulting phylogenetic tree showed high bootstrap support for a cluster of 172 sequences, which included 91 out of 92 PWID sequences collected in the 2013 survey. These patterns were consistent with an infection pattern with significant overlap or crossover of infection between the two cities. It was unlikely that the new HIV infections observed in Mandaue were introduced from somewhere outside of Cebu. Given the limited sequence lengths of our analysis, we could not definitively say that the networks were not separate, either. Prior work had shown that phylogenetic methods can detect differences in epidemics occurring across space and time [4,89-92]. It is possible that our data – outbreaks separated by two years in two adjacent cities – were at or below the limit of detection to distinguish the two as separate clusters. However, given the short genetic distanced between sequences, it was most likely that the HIV infections in the two cities were closely linked and any differences in HIV prevalence patterns arose from the differences in network structures as observed in Aim 1.

Public Health Significance

Our findings confirmed that the emergence of new HIV among PWID in Mandaue was linked to the ongoing HIV epidemic circulating among PWID in Cebu. We were unable to further delineate smaller sub-clusters of transmission, which could suggest that HIV infections were being spread across both cities, from PWID in Cebu to those in Mandaue,

and back. This suggests that the networks among PWID in the two cities were also broadly connected to each other.

Limitations

There were several limitations to the work presented here. First, we did not know the treatment status of individuals sampled. Antiretroviral therapy could suppress viral replication and alter the mutation process, which would have two consequences for our analysis. Lower viral replication would mean that PWID on ART were more likely to be missing from our data set. Among those whose HIV sequences were successfully replicated, ART could add external evolutionary pressure that would have caused the sequence to deviate from the generalized time-reversible substitution model that was used to estimate our phylogenetic trees. However, while treatment data were not directly collected, we knew through hospital registrations that fewer than 30 PWID were on ART at the time of the survey, so it would be unlikely that we captured a large number of PWID on treatment.

A second limitation was the number of sequences missing from this tree. Although all lab protocols were closely followed, we were unable to sequence over 70% of the samples in our data. Looking across a wide range of demographic and behavioral characteristics, we did not find any significant differences between PWID whose samples were successfully sequenced and those who were not; so we assume unobserved sequences were missing at random and did not impact the conclusions of this work.

Finally, analysis of the phylogenetic data was conducted on shorter consensus sequences in a relatively conserved area. Longer reads might have captured greater genetic variation across individuals, which might allow us to distinguish whether there was any separation between infections occurring in Mandaue compared to those in Cebu. However,

the increased variability along other sections of the viral sequence may have resulted in multiple substitutions that could have confounded the construction and interpretation of the phylogenetic tree. Reconstruction of phylogenetic trees when transmission order was known found that the *pol* gene, including the region that was sequenced in this analysis, was the most reliable region for phylogenetic tree reconstruction [73]. Phylogenetic analysis of the full viral population within each individual, instead of a single consensus sequence per individual, may have identified further sub-clusters of infection, but methods for such analysis are still in their infancy.

Future Research

To further assess genetic similarity and linkages between the HIV infections detected in these two cities, we might consider analysis of the complete next-generation sequencing data, which would describe the full viral population in each HIV-infected individual. Future work should sequence recently detected HIV infections, particularly those in the greater Cebu Metropolis, to assess to what degree new infections are linked to the ongoing epidemics in Cebu. The emergence of different infections could indicate other networks or communities at risk of new outbreaks or infections.

Conclusions

HIV is a serious public health concern for people who inject drugs, as new epidemics emerge rapidly and without warning. While there are widely accepted individual behaviors that put people at risk of infection, a deeper understanding of network structures and how they may expose people to infection can help in identifying and prioritizing PWID communities and individuals at greatest risk. We studied the network structures and patterns of infection among PWID in Cebu and Mandaue to better understand why we observed

differences in the introduction and spread of HIV infection among PWID in these communities.

Our results suggest that differences in the size and number of connected network components were correlated with the potential reach of infection in the two cities. We also found differences in patterns of clustering, path length, and degree, which may have helped to explain why we observed faster spread of infection in Cebu than in Mandaue. The phylogenetic results suggested that while the networks in each city exhibit different properties, the introduction of HIV into Mandaue was not through a single source. Instead, HIV infection probably crossed over from PWID in Cebu to several PWID in Mandaue through several independent transmission events. Thus, while small, fragmented network components may have offered some degree of protection from widespread infection, as long as PWID in Mandaue and other surrounding areas have contact with the high-prevalence network members (in Cebu), they will continue to be at risk of infection in the longer term.

REFERENCES

1. Skar H, Hedskog C, and Albert J. HIV-1 evolution in relation to molecular epidemiology and antiretroviral resistance. Ann N Y Acad Sci. 2011, Aug;1230108-18.

2. Niculescu I, Paraschiv S, Paraskevis D, Abagiu A, Batan I, Banica L, and Otelea D. Recent HIV-1 Outbreak Among Intravenous Drug Users in Romania: Evidence for Cocirculation of CRF14_BG and Subtype F1 Strains. AIDS Res Hum Retroviruses. 2015, May;31(5):488-95.

3. Paraskevis D, Paraschiv S, Sypsa V, Nikolopoulos G, Tsiara C, Magiorkinis G, et al. Enhanced HIV-1 surveillance using molecular epidemiology to study and monitor HIV-1 outbreaks among intravenous drug users (IDUs) in Athens and Bucharest. Infect Genet Evol. 2015, Oct;35109-21.

4. 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.

5. Redefining AIDS in Asia: Crafting an Effective Response [Internet]. Redefining AIDS in Asia: Crafting an Effective Response. 2008, Mar 26;Available from: http://data.unaids.org/pub/report/2008/20080326_report_commission_aids_en.pdf

6. 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]

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 SD, 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. AIDS Epidemic Model Impact Modeling and Analysis: Philippine Case Study (Technical Report) [Internet]. AIDS Epidemic Model Impact Modeling and Analysis: Philippine Case Study (Technical Report). 2014;[cited 2016, Jun 8] Available from: http://www.doh.gov.ph/sites/default/files/publications/2014%20AIDS%20Epidemic%20Model_Impact%20Modeling%20and%20Analysis.pdf

12. Integrated HIV Behavioral and Serological Surveillance [Internet]. Integrated HIV Behavioral and Serological Surveillance. 2013;Available from: http://www.aidsdatahub.org/sites/default/files/publication/2011_IHBSS_Comprehensive_Report.pdf

13. 2013 Integrated HIV Behavioral and Serological Surveillance (IHBSS) Report [Internet]. 2013 Integrated HIV Behavioral and Serological Surveillance (IHBSS) Report. 2014;[cited 2016, Jun 8] Available from: http://www.aidsdatahub.org/sites/default/files/publication/2013_IHBSS_Philippines_2014.pdf

14. World Drug Report [Internet]. Vienna, Austria: United Nations; 2015. [cited 2016, Feb 23] Available from: http://www.unodc.org/documents/wdr2015/World_Drug_Report_2015.pdf

15. Mathers BM, Degenhardt L, Ali H, Wiessing L, Hickman M, Mattick RP, et al. HIV prevention, treatment, and care services for people who inject drugs: a systematic review of global, regional, and national coverage [Internet]. The Lancet. 2010;375(9719):1014 - 1028.Available from: http://www.sciencedirect.com/science/article/pii/S0140673610602322

16. 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.

17. 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.

18. 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.

19. Des Jarlais DC, Perlis T, Friedman SR, Deren S, Chapman T, Sotheran JL, et al. Declining seroprevalence in a very large HIV epidemic: injecting drug users in New York City, 1991 to 1996. Am J Public Health. 1998, Dec;88(12):1801-6.

20. 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.

21. Des Jarlais DC, Friedman SR, Novick DM, Sotheran JL, Thomas P, Yancovitz SR, et al. HIV-1 infection among intravenous drug users in Manhattan, New York City, from 1977 through 1987. JAMA. 1989, Feb 17;261(7):1008-12.

22. Li H, Goggins W, and Lee SS. Multilevel analysis of HIV related risk behaviors among heroin users in a low prevalence community. BMC Public Health. 2009;9137.

23. 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.

24. 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.

25. Gupta S, Anderson RM, and May RM. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS. 1989, Dec;3(12):807-17.

26. Blower SM. HIV transmission in sexual networks: an empirical analysis. Proceedings of the Royal Society of London B: Biological Sciences. 1995;260(1359):237-244.

27. Hethcote HW, and Yorke JA. Lecture notes in biomathematics. 1984;

28. Keeling MJ. The effects of local spatial structure on epidemiological invasions. Proc Biol Sci. 1999, Apr 22;266(1421):859-67.

29. Keeling MJ, and Eames KT. Networks and epidemic models. J R Soc Interface. 2005, Sep 22;2(4):295-307.

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

31. Moore C, and Newman ME. Epidemics and percolation in small-world networks. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 2000, May;61(5 Pt B):5678-82.

32. 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.

33. Eubank S, Guclu H, Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, and Wang N. Modelling disease outbreaks in realistic urban social networks. Nature. 2004, May 13;429(6988):180-4.

34. Danon L, House TA, Read JM, and Keeling MJ. Social encounter networks: collective properties and disease transmission. J R Soc Interface. 2012, Nov 7;9(76):2826-33.

35. Newman M. Spread of epidemic disease on networks. Physical Review E. 2002, Jul;66(1):

36. Pastor-Satorras R, and Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett. 2001, Apr 2;86(14):3200-3.

37. Jones JH, and Handcock MS. Social networks: Sexual contacts and epidemic thresholds. Nature. 2003, Jun 5;423(6940):605-6; discussion 606.

38. Bansal S, Grenfell BT, and Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. J R Soc Interface. 2007, Oct 22;4(16):879-91.

39. Miller JC. Spread of infectious disease through clustered populations. J R Soc Interface. 2009, Dec 6;6(41):1121-34.

40. Ames GM, George DB, Hampson CP, Kanarek AR, McBee CD, Lockwood DR, et al. Using network properties to predict disease dynamics on human contact networks. Proc Biol Sci. 2011, Dec 7;278(1724):3544-50.

41. Bearman PS, Moody J, and Stovel K. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks1. American Journal of Sociology. 2004;110(1):44-91.

42. Potterat JJ, Rothenberg RB, and Muth SQ. Network structural dynamics and infectious disease propagation. Int J STD AIDS. 1999, Mar;10(3):182-185.

43. adams j, Moody J, and Morris M. Sex, Drugs, and Race: How Behaviors Differentially Contribute to the Sexually Transmitted Infection Risk Network Structure. American Journal of Public Health. 2013;103(2):322-329.

44. Moody J. Network Structure and Diffusion. Duke University: Durham. 2009;

45. Moody J, and Benton RA. Interdependent effects of cohesion and concurrency for epidemic potential. Ann Epidemiol. 2016, Apr;26(4):241-8.

46. Eames KT, and Keeling MJ. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proc Natl Acad Sci U S A. 2002, Oct 1;99(20):13330-5.

47. House T, and Keeling MJ. Epidemic prediction and control in clustered populations. J Theor Biol. 2011, Mar 7;272(1):1-7.

48. Latkin C, Mandell W, Vlahov D, Oziemkowska M, and Celentano D. People and places: behavioral settings and personal network characteristics as correlates of needle sharing. J Acquir Immune Defic Syndr Hum Retrovirol. 1996, Nov 1;13(3):273-80.

49. Gyarmathy VA, and Neaigus A. The effect of personal network exposure on injecting equipment sharing among IDUs in Budapest, Hungary. Connections. 2006;27(1):25-38.

50. Suh T, Mandell W, Latkin C, and Kim J. Social network characteristics and injecting HIV-risk behaviors among street injection drug users. Drug Alcohol Depend. 1997, Aug 25;47(2):137-43.

51. Paquette DM, Bryant J, and De Wit J. Use of respondent-driven sampling to enhance understanding of injecting networks: a study of people who inject drugs in Sydney, Australia. Int J Drug Policy. 2011, Jul;22(4):267-73.

52. Shaw SY, Shah L, Jolly AM, and Wylie JL. Determinants of injection drug user (IDU) syringe sharing: the relationship between availability of syringes and risk network member characteristics in Winnipeg, Canada. Addiction. 2007, Oct;102(10):1626-35.

53. De P, Cox J, Boivin JF, 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.

54. Thorpe LE, Bailey SL, Huo D, Monterroso ER, and Ouellet LJ. Injection-related risk behaviors in young urban and suburban injection drug users in Chicago (1997-1999). J Acquir Immune Defic Syndr. 2001, May 1;27(1):71-8.

55. 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.

56. Yang C, Tobin K, and Latkin C. Perceived serosorting of injection paraphernalia sharing networks among injection drug users in Baltimore, MD. AIDS Behav. 2011, Jan;15(1):16-21.

57. Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0 [Internet]. Molecular Biology and Evolution. 2013;30(12):2725-2729.Available from: http://mbe.oxfordjournals.org/content/30/12/2725.abstract

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

59. Des Jarlais DC, Perlis T, Arasteh K, Hagan H, Milliken J, Braine N, et al. "Informed altruism" and "partner restriction" in the reduction of HIV infection in injecting drug users entering detoxification treatment in New York City, 1990-2001. J Acquir Immune Defic Syndr. 2004, Feb 1;35(2):158-66.

60. 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.

61. 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.

62. Kottiri BJ, Friedman SR, Neaigus A, Curtis R, and Des Jarlais DC. Risk networks and racial/ethnic differences in the prevalence of HIV infection among injection drug users. J Acquir Immune Defic Syndr. 2002, May 1;30(1):95-104.

63. Rothenberg RB, Potterat JJ, Woodhouse DE, Darrow WW, Muth SQ, and Klovdahl AS. Choosing a centrality measure: epidemiologic correlates in the Colorado Springs study of social networks. Social Networks. 1995;17(3):273-297.

64. 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.

65. Hoffmann JP, Su SS, and Pach A. Changes in network characteristics and HIV risk behavior among injection drug users. Drug Alcohol Depend. 1997, Jun 6;46(1-2):41-51.

66. 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.

67. Friedman SR, Kottiri BJ, Neaigus A, Curtis R, Vermund SH, and Des Jarlais DC. Network-related mechanisms may help explain long-term HIV-1 seroprevalence levels that remain high but do not approach population-group saturation. Am J Epidemiol. 2000, Nov 15;152(10):913-22.

68. Wylie JL, Shah L, and Jolly AM. Demographic, risk behaviour and personal network variables associated with prevalent hepatitis C, hepatitis B, and HIV infection in injection drug users in Winnipeg, Canada. BMC Public Health. 2006;6229.

69. Logan JJ, Jolly AM, and Blanford JI. The Sociospatial Network: Risk and the Role of Place in the Transmission of Infectious Diseases. PLoS One. 2016;11(2):e0146915.

70. Havens JR, Lofwall MR, Frost SD, Oser CB, Leukefeld CG, and Crosby RA. Individual and network factors associated with prevalent hepatitis C infection among rural Appalachian injection drug users. Am J Public Health. 2013, Jan;103(1):e44-52.

71. Young AM, Jonas AB, and Havens JR. Social networks and HCV viraemia in anti-HCV-positive rural drug users. Epidemiol Infect. 2012, Apr 19;1-10.

72. 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.

73. Lemey P, and Vandamme AM. Exploring full-genome sequences for phylogenetic support of HIV-1 transmission events (letter). AIDS. 2005, Sep 23;19(14):1551-2.

74. 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.

75. Sacks-Davis R, Daraganova G, Aitken C, Higgs P, Tracy L, Bowden S, et al. Hepatitis C virus phylogenetic clustering is associated with the social-injecting network in a cohort of people who inject drugs. PLoS One. 2012;7(10):e47335.

76. Aitken CK, McCaw RF, Bowden DS, Tracy SL, Kelsall JG, Higgs PG, et al. Molecular epidemiology of hepatitis C virus in a social network of injection drug users. J Infect Dis. 2004, Nov 1;190(9):1586-95.

77. Paintsil E, Verevochkin SV, Dukhovlinova E, Niccolai L, Barbour R, White E, et al. Hepatitis C virus infection among drug injectors in St Petersburg, Russia: social and molecular epidemiology of an endemic infection. Addiction. 2009, Nov;104(11):1881-90.

78. 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.

79. Liu J, and Zhang C. Phylogeographic analyses reveal a crucial role of Xinjiang in HIV-1 CRF07_BC and HCV 3a transmissions in Asia. PLoS One. 2011;6(8):e23347.

80. Mullick R, Sengupta S, Sarkar K, Saha MK, and Chakrabarti S. Phylogenetic analysis of env, gag, and tat genes of HIV type 1 detected among the injecting drug users in West Bengal, India. AIDS Res Hum Retroviruses. 2006, Dec;22(12):1293-9.

81. 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.

82. Ciccozzi M, Montieri S, Salemi M, De Oliveira T, Dorrucci M, Sinicco A, et al. An outbreak of HIV-1 subtype G among Italian injecting drug users. AIDS. 2007, May 31;21(9):1213-5.

83. Lukashov VV, Jurriaans S, Bakker M, and Berkhout B. Transmission of risk-group specific HIV-1 strains among Dutch drug users for more than 20 years and their replacement by nonspecific strains after switching to low-harm drug practices. J Acquir Immune Defic Syndr. 2013, Feb 1;62(2):234-8.

84. Peters PJ, Pontones P, Hoover KW, Patel MR, Galang RR, Shields J, et al. HIV Infection Linked to Injection Use of Oxymorphone in Indiana, 2014-2015. N Engl J Med. 2016, Jul 21;375(3):229-39.

85. Ferdats A, Konicheva V, Dievberna I, Lilja E, and Albert J. An HIV type 1 subtype A outbreak among injecting drug users in Latvia. AIDS Res Hum Retroviruses. 1999, Nov 1;15(16):1487-90.

86. Nabatov AA, Kravchenko ON, Lyulchuk MG, Shcherbinskaya AM, and Lukashov VV. Simultaneous introduction of HIV type 1 subtype A and B viruses into injecting drug users in southern Ukraine at the beginning of the epidemic in the former Soviet Union. AIDS Res Hum Retroviruses. 2002, Aug 10;18(12):891-5.

87. Lin YT, Lan YC, Chen YJ, Huang YH, Lee CM, Liu TT, et al. Molecular epidemiology of HIV-1 infection and full-length genomic analysis of circulating recombinant form 07_BC strains from injection drug users in Taiwan. J Infect Dis. 2007, May 1;195(9):1283-93.

88. Tovanabutra S, Watanaveeradej V, Viputtikul K, De Souza M, Razak MH, Suriyanon V, et al. A new circulating recombinant form, CRF15_01B, reinforces the linkage between IDU and heterosexual epidemics in Thailand. AIDS Res Hum Retroviruses. 2003, Jul;19(7):561-7.

89. 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.

90. Tovanabutra S, Beyrer C, Sakkhachornphop S, Razak MH, Ramos GL, Vongchak T, et al. The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002. AIDS Res Hum Retroviruses. 2004, May;20(5):465-75.

91. Op de Coul EL, Prins M, Cornelissen M, van der Schoot A, Boufassa F, Brettle RP, et al. Using phylogenetic analysis to trace HIV-1 migration among western European injecting drug users seroconverting from 1984 to 1997. AIDS. 2001, Jan 26;15(2):257-66.

92. Skar H, Axelsson M, Berggren I, Thalme A, Gyllensten K, Liitsola K, et al. Dynamics of two separate but linked HIV-1 CRF01_AE outbreaks among injection drug users in Stockholm, Sweden, and Helsinki, Finland. J Virol. 2011, Jan;85(1):510-8.

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

94. Heckathorn DD. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. Social Problems. 2002, Feb;49(1):11-34.

95. Gile KJ. Improved Inference for Respondent-Driven Sampling Data With Application to HIV Prevalence Estimation. Journal of the American Statistical Association. 2012, Jan;106(493):135-146.

96. Gile KJ, Johnston LG, and Salganik MJ. Diagnostics for respondent-driven sampling. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2015;178(1):241-269.

97. Volz E, and Heckathorn DD. Probability based estimation theory for respondent driven sampling. Journal of official statistics. 2008;24(1):79.

98. Lu X. Linked Ego Networks: Improving estimate reliability and validity with respondent-driven sampling. Social Networks. 2013, Oct;35(4):669-685.

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

100. Merli MG, Moody J, Smith J, Li J, Weir S, and Chen X. Challenges to recruiting population representative samples of female sex workers in China using Respondent Driven Sampling. Social Science & Medicine. 2015, Jan;12579-93.

101. Snijders TA. Statistical Models for Social Networks. Annual Review of Sociology. 2011, Aug 11;37(1):131-153.

102. Goodreau SM, Kitts JA, and Morris M. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks*. Demography. 2009;46(1):103-125.

103. 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.

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

105. Rolls DA, Wang P, Jenkinson R, Pattison PE, Robins GL, Sacks-Davis R, et al. Modelling a disease-relevant contact network of people who inject drugs. Social Networks. 2013, Oct;35(4):699-710.

106. Viger F, and Latapy M. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. Journal of Complex Networks. 2016, Mar;4(1):15-37.

107. Valente TW. Social Networks and Health: Models, Methods, and Applications [Internet]. New York: Oxford University Press; 2010. [cited 2013, Dec 3] Available from: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195301014.001.0001/acprof-9780195301014

108. Wasserman S, and Faust K. Social network analysis: Methods and applications. Cambridge: Cambridge University Press; 1994.

109. 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.

110. Verdery AM, Merli MG, Moody J, Smith JA, and Fisher JC. Respondent-driven samplign estimators under real and theoretical recruitment conditions of female sex workers in China. Epidemiology. 2015, Sep;26(5):661-665.

111. Rambaut A, Posada D, Crandall KA, and Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet. 2004, Jan;5(1):52-61.

112. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004, Aug 19;5113.

113. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7.

114. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014, May 1;30(9):1312-3.

115. Hué S, Pillay D, Clewley JP, and Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc Natl Acad Sci U S A. 2005, Mar 22;102(12):4425-9.

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

117. Global HIV Statistics [Internet]. UNAIDS Fact Sheet. [cited 2016, Dec 1] Available from: http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

118. Bosch X. HIV mystery in the Philippines. Lancet Infect Dis. 2003, Jun;3(6):320.

119. HIV/AIDS and ART Registry of the Philippines (HARP) [Internet]. HIV/AIDS and ART Registry of the Philippines (HARP) . 2016, Mar;Available from: http://www.doh.gov.ph/sites/default/files/statistics/EB_HIV_Mar-AIDSreg2016.pdf

120. DE. Philippine Estimates of Most-at-Risk Population and People Living with HIV. Philippine Estimates of Most-at-Risk Population and People Living with HIV. 2011;

121. R Core Team. R: A language and environment for statistical computing [Internet]. R: A language and environment for statistical computing. 2013;Available from: http://www.R-project.org/

122. Handcock S, Fellows E, and Gile J. RDS: Respondent-Driven Sampling, Version 0.7-6 [Internet]. RDS: Respondent-Driven Sampling, Version 0.7-6. 2012;[cited 2015] Available from: http://hpmrg.org

123. Csardi G, and Nepusz T. The igraph software package for complex network research [Internet]. The igraph software package for complex network research. 2006;Complex Systems1695.Available from: http://igraph.org

124. Handcock M, Hunter D, Butts C, Goodreau S, Krivitsky P, and Morris M. statnet: Software tools for the statistical analysis of network data [Internet]. statnet: Software tools for the statistical analysis of network data . 2016;[cited 2013, Dec 18] Available from: http://www.statnet.org/

125. Handcock MS, Hunter DR, Butts CT, Goodreau SM, and Morris M. statnet: Software tools for the representation, visualization, analysis and simulation of network data [Internet]. Journal of Statistical Software. 2008;24(1):1-11.Available from: http://www.jstatsoft.org/v24/i01

126. Rothenberg RB, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, and Klovdahl AS. Social network dynamics and HIV transmission. AIDS. 1998, Aug 20;12(12):1529-36.

127. Gyarmathy VA, Li N, Tobin KE, Hoffman IF, Sokolov N, Levchenko J, et al. Injecting equipment sharing in Russian drug injecting dyads. AIDS Behav. 2010, Feb;14(1):141-51.

128. Dukhovlinova E, Masharsky A, Toussova O, Verevochkin S, Solovyeva T, Meringof M, et al. Two Independent HIV Epidemics in Saint Petersburg, Russia Revealed by Molecular Epidemiology. AIDS Research and Human Retroviruses. 2015, Jun;31(6):608-614.

129. Johnston L, Saumtally A, Corceal S, Mahadoo I, and Oodally F. High HIV and hepatitis C prevalence amongst injecting drug users in Mauritius: findings from a population size estimation and respondent driven sampling survey. Int J Drug Policy. 2011, Jul;22(4):252-8.

130. (UNAIDS). The gap report. Geneva: UNAIDS; 2014.

131. Mirzoyan L, Berendes S, Jeffery C, Thomson J, Ben Othman H, Danon L, et al. New evidence on the HIV epidemic in Libya: why countries must implement prevention programs among people who inject drugs. J Acquir Immune Defic Syndr. 2013, Apr 15;62(5):577-83.

132. Global State of Harm Reduction 2014 [Internet]. London, UK: Harm Reduction International; 2014. Available from: http://www.ihra.net/files/2015/02/16/GSHR2014.pdf

133. Telan EF, Samonte GM, Abellanosa-Tac-An IP, Alesna ET, Leaño PS, Emphasis YE, 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.

134. Telan EF, Samonte GM, Palaypayon N, Abellanosa-Tac-An IP, Leaño PS, 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.

135. Salganik MJ, and Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. Sociological methodology : an official publication of the American Sociological Association. 2004;34193-.

136. Rao JN, and Scott AJ. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. The Annals of Statistics. 1984;46-60.

137. 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.

138. Gile KJ, and Handcock MS. Network model-assisted inference from respondent-driven sampling data. J. R. Stat. Soc. A. 2015, Jan 27;178(3):619-639.

139. Handcock MS, and Gile KJ. COMMENT: ON THE CONCEPT OF SNOWBALL SAMPLING. Sociological Methodology. 2011, Aug;41(1):367-371.

140. Latkin CA, Kuramoto SJ, Davey-Rothwell MA, and Tobin KE. Social norms, social networks, and HIV risk behavior among injection drug users. AIDS Behav. 2010, Oct;14(5):1159-68.

141. 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.

142. Helleringer S, and Kohler HP. Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi. AIDS. 2007, Nov 12;21(17):2323-32.

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