Social network clustering and the spread of HIV/AIDS among persons who inject drugs in two cities in the Philippines

Ashton M. Verdery, M.A., Ph.D.[†,1], **Nalyn Siripong, M.Sc., Ph.D.**[2], and **Brian W. Pence, M.P.H., Ph.D.**[2]

[1]Department of Sociology and Criminology, Population Research Institute, and Institute for CyberScience, The Pennsylvania State University, University Park, PA

[2]Department of Epidemiology, Gillings School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract

INTRODUCTION—The Philippines has seen rapid increases in HIV prevalence among people who inject drugs. We study two neighboring cities where a linked HIV epidemic differed in timing of onset and levels of prevalence. In Cebu, prevalence rose rapidly from under 1% to 54% between 2009 and 2011 and remained high through 2013. In nearby Mandaue, HIV remained below 4% through 2011 then rose rapidly to 38% by 2013.

OBJECTIVES—We hypothesize that infection prevalence differences in these cities may owe to aspects of social network structure, specifically levels of network clustering. Building on prior research, we hypothesize that higher levels of network clustering are associated with greater epidemic potential.

METHODS—Data were collected with respondent-driven sampling among males who inject drugs in Cebu and Mandaue in 2013. We first examine sample composition using estimators for population means. We then apply new estimators of network clustering in respondent-driven sampling data to examine associations with HIV prevalence.

RESULTS—Samples in both cities were comparable in terms of composition by age, education, and injection locations. Dyadic needle sharing levels were also similar between the two cities, but network clustering in the needle sharing network differed dramatically. We found higher clustering in Cebu than Mandaue, consistent with expectations that higher clustering is associated with faster epidemic spread.

CONCLUSION—This paper is the first to apply estimators of network clustering to empirical respondent-driven samples, and it offers suggestive evidence that researchers should pay greater attention to network structure's role in HIV transmission dynamics.

Keywords

Respondent-driven sampling; Social Networks; HIV/AIDS; Clustering Coefficients; PWID; The Philippines

[†]corresponding author: amv5430@psu.edu.

Introduction

The Philippines HIV epidemic was small until 2010, with prevalence remaining low even among high risk populations [1]. Current estimates imply that between 21,000 and 100,000 individuals in the Philippines are living with HIV [2], but recent trends suggest a larger epidemic could take hold. From 2000 to 2012 new infections more than doubled, which contrasts markedly with large declines in other Southeast Asian countries [3]. New infections in the Philippines are concentrated among high risk populations in large cities, especially people who inject drugs (PWID), female sex workers, and men who have sex with men. Of these groups, male PWID have experienced the starkest change, with measured prevalence rising from less than 1% to 54% between 2009 and 2011 in the major city of Cebu, where it has remained steady since [4–6]. Other high risk groups saw smaller changes, with MSM reaching the second highest prevalence rates at 4% [5].

The explosion of HIV among male PWID in the Philippines presents an intriguing puzzle. Mandaue is a city about a third the size of Cebu and is approximately a 25 minute drive away. Despite their proximity, Mandaue did not experience increases in HIV infection for several years after the large increases in Cebu. In 2011, when HIV prevalence among male PWID had already reached its peak in Cebu, it remained at 4% among Mandaue's male PWID [5]. However, by 2013 HIV among male PWID had also spread in Mandaue, with prevalence reaching 38% [5]. Estimated prevalence of 5% from 2015 data suggest a dip in prevalence in Mandaue or else that the 2013 prevalence estimates were too high [6]. In either case, Mandaue's delayed, slower, or more marginal increase in prevalence may be a result of later introduction of infection into the population, but network structures may also have played a role in limiting the spread of HIV in this city.

In this paper, we explore how risk networks might have slowed the spread of HIV among male PWID in Mandaue compared to Cebu. Individuals embedded in different network structures have different risk environments that can alter the dynamics of HIV outbreaks [7,8]. For instance, a city containing small clusters of PWID who only share needles with one another may be at lower risk of epidemic expansion than one where PWID are more diffuse in their needle sharing behaviors. We use new methods for estimating structural properties of social networks from respondent-driven sampling (RDS) data [9–11] to assess whether differences in the injection risk networks in Cebu and Mandaue may explain differences in their rates of HIV spread. Network clustering, the tendency for one's friends to be friends with one another, is a particularly important measure in the social networks literature that alters the speed and ultimate size of epidemic expansion [12,13]. We focus on the role of network clustering in needle sharing behaviors, the tendency for a person's needle sharing partners to share needles with one another. Other work finds that network clustering is a significant feature of co-injection networks [14]. We find substantially higher clustering in Cebu than Mandaue, which is consistent with hypotheses from network diffusion models that predict that higher clustering is associated with more rapid epidemic spreading and lower epidemic thresholds. This paper is the first application to examine network clustering in RDS samples and highlights a) the role that the structural properties of social networks might play in advancing HIV epidemics, and b) the potential for RDS data to allow researchers to measure such structural properties.

Background

Respondent-driven sampling

PWID and other groups at high risk of HIV are hard to survey. Such groups are often called hidden populations [11] that traditional probability designs have difficulty surveying because they lack a sampling frame or are members of stigmatized groups that avoid survey participation [15]. RDS is an innovative approach to studying hidden populations that capitalizes on social networks in the target population by asking and incentivizing respondents to recruit peers into the survey. Under certain conditions, data obtained with RDS can provide probability-based inferences about univariate statistics like infection prevalence [16–20]. The possibility of rapidly, cost-effectively, and anonymously sampling hidden populations, with the potential promise of obtaining generalizable estimates, has popularized RDS and led to its use in hundreds of studies [21–23].

RDS begins with a convenience sample of "seeds" selected from the target population. Researchers ask each seed to recruit two other members of the hidden population that they know who have not already participated [24]. Subsequent participants are also asked to recruit others in "waves", where recruitment waves are designated as number of steps along a recruitment "chain" starting a seed. This process repeats until the desired sample size is attained or respondents can no longer recruit. RDS studies typically use a "dual incentive" structure that compensates respondents for their own participation ("primary incentive") and offers additional "secondary" incentives that respondents can collect after successful recruitment, allowing for the prospect of a second interview ("follow-up survey"). By asking participants themselves to identify and recruit others, RDS data collection methods facilitate both recruitment and anonymous data collection.

RDS is most commonly used to estimate HIV prevalence or risk factors through the application of estimators designed to account for its tendency to oversample popular members [16,19,25]. However, data collected as part of an RDS survey contain additional information that can provide valuable insights into social network structure [26]. For instance, RDS asks each respondent about his personal network size. Some RDS surveys also examine the extent to which respondents resemble their recruiters [27,28], which is similar to the notion of homophily in sociology [29], though this practice is controversial [30].

Clustering Coefficients

An additional network measure that can be obtained from RDS surveys [9,10] and may be useful for studying the spread of infectious disease is the clustering coefficient [12,13,31]. For individuals in friendship networks, we can say that each person's level of network clustering is the proportion of his friends who are friends with each other. This feature is measured using network "triplets", or unique groups created from sets of three distinct people i, j, and k, where at a minimum i has a social network tie with both j and k. Triplets can either be open, where there is no tie between j and k, or closed (also called "triangles"), where there is a tie between j and k.

The number of triplets, each of which can either be open or closed, that person i is a member of is determined by his network size, d_i: *number of triplets for $i = d_i(d_i - 1)/2$*. For example, Figure 1 shows a hypothetical network of PWID linked by co-injecting ties (solid lines) and sampled with RDS (dashed arrows). For now, we concentrate on the co-injecting ties. In this network, Person A co-injects with 3 people (B, C, and D), making him a member of 3 co-injecting triplets (ABC, ACD, ABD), but only one of those co-injecting triplets is closed (ACD), so this person's co-injecting network clustering is 1/3. By contrast, from his perspective, person C is only a member of a closed co-injecting triplet, so he has co-injecting network clustering of 1.0. From each of their perspectives, persons B and F are only members of open triplets and as such have network clustering of 0.0. Person E is not involved in any co-injecting triplets so his clustering level is also 0.0. At the network level, researchers have designed several statistics to summarize a network's tendency toward clustering [9]. Two that we focus on are known as the global clustering coefficient, which sums the number of observed closed triplets in the network and divides by the total number of triplets (closed or open), and the local clustering coefficient, which takes the average of each person's network clustering [10].

Clustering is a hallmark of human social networks. People have ties with the friends of their friends more commonly than would be expected by chance. Clustering is a key criterion to determine whether a social network has "small-world" properties, which can be distinguished by high levels of clustering and low average distances between network members [12]. Small world networks are consistent with rapid infection spread because people in small world networks tend to have few degrees of separation between them and multiple pathways for transmission [32]. Holding the distribution of personal network sizes, infectivity, and the timing and number of introduction points constant, increases in clustering in a co-injection network from a low level can be expected to speed the rate of transmission and the size of the resultant epidemic because there are more paths between co-injecting partners [13].

Based on the observed spread of HIV among PWID in Cebu and Mandaue and theoretical expectations from network diffusion models of the effects of network clustering on diffusion of infections, we hypothesize that the needle sharing network in Cebu had a higher clustering coefficient than the needle sharing network in Mandaue because it had a much more rapid spread of HIV among its PWID. Under our hypothesis, the increased diffusive capacity of networks with more clustering would constitute the mechanism by which HIV spread more quickly in Cebu than Mandaue, which implies that we expect that the PWID populations in these two cities will be similar on other critical features relevant to HIV transmission such as injection frequency, dyadic needle sharing, and population composition.

Data and Methods

We use data collected with RDS among male PWID in 2013 in Cebu and Mandaue. In each city, local health officers collaborated with peer educators to identify seven seed PWID and standard RDS procedures were used. Researchers offered participants a primary incentive of 200 Philippine Pesos (US $4) for completing the interview and blood sample and a

secondary incentive of 50 Pesos for each successful recruitment, for a total maximum of 300 PHP (approximately US $6). When a PWID retrieved his secondary incentive, study staff administered a brief follow-up survey about aspects of recruitment and the relationships between recruiters and recruitees. Data collection was a surveillance activity and was not subject to institutional review board (IRB) approval. Data analysis received IRB approval from *[blinded for anonymity].*

Figure 1 also displays a hypothetical RDS sample on the population, as well as respondent reports on the existence of co-injecting ties (hashes). This sample started from node A, who recruited B and C into the survey. Node C then recruited D, who recruited E, who recruited F. During secondary interviews, survey respondents were asked about co-injecting behaviors. Specifically, they were asked "(Q1): What is your relationship with the person who accepted your coupon?" and "(Q2): What is the relationship between your recruiter and the person who accepted your coupon?" For both questions, respondents could choose all that apply from a set of choices that included "shares needles with this person" and "pools drugs with this person," both of which we consider to indicate that the relevant parties co-inject. Using these reports, we can define a subset of the co-injecting ties shown in Figure 1 that would be observed during an RDS survey. Co-injecting ties in Figure 1 are marked with hashes if reported in response to the first question and double hashes if reported in response to the second question. For instance, during node A's follow-up interview, he reports on the existence of co-injecting ties between himself and his recruits B and C. Node B did not recruit anyone so he does not return for a follow-up interview and does not report on co-injecting ties. When node C returns, he reports the existence of a co-injecting tie with node D in response to Q1 and the existence of a co-injecting tie between his recruiter A and his recruitee D in response to Q2. Node D reports that he does not co-inject with his recruitee E during Q1 and that E and C do not co-inject during Q2. When node E returns he would report that he does not co-inject with his recruitee F but that F co-injects with D.

We estimate clustering coefficients among the male PWID populations sampled with RDS in Cebu and Mandaue using modified estimators, originally defined for internet samples [10] and recently expanded for in-person RDS studies [9]. These estimators work by defining recruiter-recruit-recruitee triplets in the RDS sample and taking a degree weighted average of the number that are closed. In Figure 1, the set of recruiter-recruit-recruitee triplets observed in the RDS sample is ACD, CDE, and DEF. In these triplets, seed respondents can only be the recruiter, while those who do not recruit can only be recruitees. We code triplets as closed or open on the basis of responses during the secondary interview to the co-injecting questions described above. We consider a co-injecting triplet closed if three conditions are met: 1) the respondent's recruiter reported sharing needles with the respondent, 2) the respondent reported sharing needles with his recruitee, and 3) the respondent reported that his recruitee shares needles with his recruiter. We code all other cases as open co-injecting triplets. Thus, in Figure 1, we code the triplet ACD as closed because A reported sharing needles with C, and C reported sharing with D and that A and D also co-inject. By contrast, we code the triplet CDE as open because D reported that C and E do not co-inject. We also code the triple DEF as open because E does not co-inject with D or F. In Appendix A, we consider three alternate triplet definitions; all approaches yield results consistent with our conclusions.

We first examine univariate statistics for all respondents to the primary survey using RDS estimators of population means and confidence intervals [17]. We then examine network clustering in the co-injecting network using the closed and open triplet definitions described above. For this, we apply estimators for network clustering with RDS data that are robust to the features of realistic RDS surveys, including biased recruitment and co-injection reporting error [9]. Unfortunately, methods for estimating confidence intervals around point estimates of clustering in RDS data have not been developed. However, unlike RDS mean estimators [20,33], RDS clustering estimators tend to exhibit low sampling variance, which suggests that substantial differences between samples reflect differences in population parameters rather than sampling error [9].

Results

Target sample sizes were reached in both cities, with 457 recruits and seeds in Cebu and 310 recruits and seeds in Mandaue. One seed from each city failed to initiate recruitment, leaving 6 analyzable RDS chains in each city. Cebu had the largest recruitment chain, where one seed yielded 214 recruits across 12 waves. Mandaue had the longest recruitment chain from a seed, recruiting 87 PWID over 21 waves of recruitment (Table 1). We focus on the 753 individuals who were recruited into the survey (i.e., non-seeds). Because of their complexity, we review sample size and missing data issues in Appendix B.

General demographic characteristics and injecting behaviors were similar across both cities (Table 2). PWID in Cebu were slightly older than those in Mandaue, although a majority in both were under 35 years of age. In both cities, over 40% of PWID completed elementary school and one-third completed at least high school. Most PWID in either sample visited shooting galleries to inject drugs. One-third of PWID in each of the two cities began injecting before age 18, and, on average, PWID across both cities have been injecting for around 10 years. Responses to several questions about needle sharing at last injection and over the past 6 months are also similar between the cities. As noted before, HIV prevalence in the Cebu sample was higher than it was in the Mandaue sample (52% vs. 34%; note these numbers differ from surveillance reports because we apply different estimators).

Data on network size were collected on the PWID friendship network with a series of four questions following best practices in RDS [34,35]. Reported PWID friendship network size was higher in Cebu, where respondents knew more than 20 other PWID on average, compared to Mandaue, where they knew only 8 others on average. Though the network size questions in the survey ask only about social networks, and not co-injecting networks, several questions in the survey examined dyadic levels of co-injection within a subset of respondents' overall social networks. Responses to these questions indicate that over 75% of respondents in both cities reported injecting with their recruiter, over half of whom did so quite regularly.

Of the 753 recruits in the two cities, 342 were non-recruiters. Non-recruiters were ineligible for secondary incentives and did not complete follow-up interviews. Additionally, 47 recruiters did not return for secondary incentives or complete follow-up interviews. We could not analyze clustering for seeds, non-recruiters, and those who did not complete the

follow-up survey. Recent work shows that excluding cases without follow-up can increase biases and error in RDS clustering coefficient estimators, though error rates remain low [9]. We examine clustering in the co-injection network for the remaining 364 participants; 207 in Cebu and 157 in Mandaue. In total, we collected data on 656 triplets along the recruitment chains, 374 in Cebu and 282 in Mandaue.

Because not all respondents answered whether their recruiters share needles with their recruitees, we were unable to code all triplets as either open or closed according to the definitions above. In Cebu, there were missing data on these questions for 7.2% of the 374 triplets. In Mandaue, data were missing for 2.2% of the 282 triplets. Though we believe that non-response on these questions is consistent with an answer of non-sharing, we evaluate sensitivity to alternate approaches to dealing with these missing data. We first compute clustering coefficients assuming triplets are open for missing cases, then we assume they are closed, and finally we drop all triplets with missing responses. We reach the same conclusions from all three approaches.

Table 3 shows both measures of the clustering coefficients computed under all three assumptions in the two cities. For both measures, under all three assumptions, we find higher clustering coefficients in Cebu than in Mandaue. Methodological work extending clustering coefficient estimation to RDS has indicated that the local clustering coefficient tends to provide more consistent results than the global clustering coefficient in realistic RDS samples [9]. In our results, both then local and global clustering coefficient measures yield the same conclusions.

Discussion

HIV spread rapidly among male PWID in the Philippines. Examining changes in prevalence among the 269 subnational PWID epidemics documented by the Joint United Nations Programme on HIV/AIDS [36], we find only 13 (5%) with quicker expansion than that seen in Cebu from 2009–2011. This leads us to ask: What about the PWID population in Cebu led to such rapid HIV expansion? We considered this question with RDS data drawn from the hidden populations of PWID in Cebu and nearby Mandaue and used new methods to test whether the structure of the co-injection networks in these two cities might differ in such a way as to speed spread in Cebu and slow it in Mandaue.

The RDS samples drawn among male PWID in Cebu and Mandaue were comparable in terms of age, education, and preferred locations to inject and procure injecting equipment. They also had very similar levels of dyadic co-injecting between recruiters and recruits. One substantial difference between these populations, however, was seen in the levels of clustering in their co-injection networks. Based on theoretical models of network diffusion, the levels of clustering we observed were consistent with a more rapid outbreak of HIV/AIDS in Cebu than in Mandaue. Higher levels of clustering are associated with a lower epidemic threshold in mathematical models [13], meaning that increases in clustering should accelerate the speed of epidemic spread for infections like HIV [37]. An alternative theory is that networks with high levels of clustering have greater skew in the degree distribution, which also accelerates epidemic spread [38,8,39], but we cannot disentangle these

mechanisms because the data we examine lack reliable measures of co-injecting degree distributions. We find much higher levels of clustering in Cebu than Mandaue, which aligns with the historical dynamics of the HIV epidemics in these cities.

Considering the changes in HIV prevalence in Cebu and Mandaue over the period 2007–2015, there is another puzzle: why does HIV prevalence appear to have peaked in Cebu? We believe network theory also offers guidance here. Higher levels of clustering are associated with reductions in the ultimate extent of epidemic spread because increases in clustering reduce the size of the largest group of people connected through chains of relations (i.e., the giant component) [13]. A similar idea holds that HIV can exhibit "network firewall dynamics," where infected individuals who have passed high transmissibility periods prevent epidemic saturation by effectively disconnecting the network [40].

Our analysis is subject to several limitations. Unfortunately, earlier surveillance efforts in these populations did not include follow-up interviews that asked the network clustering questions, which precludes a longitudinal analysis of associations between changes in clustering and prevalence. A second limitation is that the surveys we examine used a binary measure of clustering, while methodological work suggests that asking respondents how many of their potential recruitees are known by, or co-inject with, their recruiter can yield more accurate and robust results [9]. A third limitation of this study is that RDS has high sampling variance for prevalence estimates [33,41]. Indeed, estimates of HIV prevalence among PWID in Mandaue show such variability. However, prevalence estimates for Cebu were consistently high across multiple survey waves, which reinforces our confidence in them. Further, even if the 2013 prevalence estimate for Mandaue is too high, such results would be consistent with our hypotheses about the role of network clustering in slowing HIV expansion in this city compared to Cebu. It is worth reiterating that RDS clustering coefficient estimates tend to have much lower sampling variance than RDS prevalence estimates [9].

Our study has implications for HIV research and interventions among PWID populations. We follow Dombrowski et al. [14] and argue for improved data collection approaches that can reveal the role of structural risk factors in epidemic expansion. Absent such data, it will be challenging for researchers to develop effective network interventions to reduce harm among PWID populations [42]. Because our results suggest little difference between Cebu and Mandaue in terms of the proportion of users who tend to co-inject drugs, interventions targeting co-injection levels may be less effective than one targeting co-injection practices (e.g., co-injection network clustering). We leave this question for future work.

Finally, we note that the changing political context in the Philippines may present particular challenges for continued research among the key PWID population and may exacerbate the spread of HIV. In addition to its direct human costs, the further stigmatization of drug use has the potential to degrade surveillance efforts and may prevent respondents from seeking treatment and encourage the increases in unsafe injection practices.

Conclusion

This paper is the first application of network clustering measures to data collected with RDS, and it extends broad considerations in the social networks literature about how network clustering relates to the transmission of infections and health behaviors. We study cross-sectional RDS surveys of male PWID collected in 2013 from two neighboring cities in the Philippines that have experienced varied HIV trajectories. We posited that aspects of network structure may be related to these trends, and hypothesized that network clustering among a co-injecting network would exhibit a positive relationship with HIV prevalence. We found clustering levels in these cities that accord with our expectations such that the city with the lower levels of clustering (Mandaue) also had lower HIV prevalence and a history that suggests less rapid or more marginal expansion of HIV than occurred in the city with higher clustering (Cebu). This paper offers a suggestive research agenda that brings consideration of network structure in RDS samples to the forefront, and proposes that researchers should investigate longitudinal changes in network clustering and their association with HIV prevalence trends to more fully explore this relationship.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Farr AC, Wilson DP. An HIV epidemic is ready to emerge in the Philippines. J. Int. AIDS Soc. 2010; 13:16. [PubMed: 20409346]

2. UNAIDS. [cited 2016 Jun 12] Philippines | UNAIDS [Internet]. 2014. Available from: http://www.unaids.org/en/regionscountries/countries/philippines

3. Phanuphak N, Lo Y-R, Shao Y, Solomon SS, O'Connell RJ, Tovanabutra S, et al. HIV Epidemic in Asia: Implications for HIV Vaccine and Other Prevention Trials. AIDS Res. Hum. Retroviruses. 2015; 31:1060–76. [PubMed: 26107771]

4. UNAIDS. [cited 2016 Jun 12] 2014 Global AIDS Response Progress Reporting, Country Progress Report, Philippines [Internet]. 2014. Available from: http://www.unaids.org/sites/default/files/country/documents//PHL_narrative_report_2014.pdf

5. Department of Health, Republic of the Philippines. AIDS Epidemic Model Impact Modeling & Analysis Philippine Case Study Technical Report 2014 [Internet]. 2014. Available from: http://www.doh.gov.ph/node/5452

6. National Epidemiology Center, Department of Health, Philippines. [cited 2017 Apr 14] 2015 Integrated HIV Behavioral and Serologic Surveillance (IHBSS). [Internet]. 2015. Available from: http://www.aidsdatahub.org/sites/default/files/highlight-reference/document/Philippines_2015_IHBSS_Factsheets.pdf

7. Bearman PS, Moody J, Stovel K. Chains of affection: The structure of adolescent romantic and sexual networks1. Am. J. Sociol. 2004; 110:44–91.

8. Dombrowski K, Curtis R, Friedman S, Khan B. Topological and Historical Considerations for Infectious Disease Transmission among Injecting Drug Users in Bushwick, Brooklyn (USA). World J. AIDS. 2013; 3:1–9. [PubMed: 24672745]

9. Verdery AM, Fisher JC, Siripong N, Abdesselam K, Bauldry S. New Survey Questions and Estimators for Estimating Network Clustering with Respondent-driven Sampling Data. Sociol. Methodol. [Internet]. 2017 Available from: osf.io/w3d86.

10. Hardiman, SJ., Katzir, L. Estimating clustering coefficients and size of social networks via random walk. Proc. 22nd Int. Conf. World Wide Web; International World Wide Web Conferences Steering Committee; 2013. p. 539-50.

11. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc. Probl. 1997; 44:174–99.

12. Watts DJ, Strogatz SH. Collective dynamics of "small-world"networks. nature. 1998; 393:440–2. [PubMed: 9623998]

13. Newman ME. Properties of highly clustered networks. Phys. Rev. E. 2003; 68:026121.

14. Dombrowski K, Khan B, McLean K, Curtis R, Wendel T, Misshula E, et al. A Reexamination of Connectivity Trends via Exponential Random Graph Modeling in Two IDU Risk Networks. Subst. Use Misuse. 2013; 48:1485–97. [PubMed: 23819740]

15. Kalton G. Methods for oversampling rare subpopulations in social surveys. Surv. Methodol. 2009; 35:125–41.

16. Volz E, Heckathorn DD. Probability based estimation theory for respondent driven sampling. J. Off. Stat. 2008; 24:79.

17. Gile KJ. Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. J. Am. Stat. Assoc. 2011; 106

18. Lu X. Linked Ego Networks: Improving estimate reliability and validity with respondent-driven sampling. Soc. Netw. 2013; 35:669–85.

19. Verdery AM, Merli MG, Moody J, Smith JA, Fisher JC. Respondent-driven Sampling Estimators Under Real and Theoretical Recruitment Conditions of Female Sex Workers in China. Epidemiology. 2015; 26:661–5. [PubMed: 26214337]

20. Verdery AM, Mouw T, Bauldry S, Mucha PJ. Network Structure and Biased Variance Estimation in Respondent Driven Sampling. PLoS ONE. 2015; 10:e0145296. [PubMed: 26679927]

21. Malekinejad M, Johnston LG, Kendall C, Kerr LRFS, Rifkin MR, Rutherford GW. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. AIDS Behav. 2008; 12:105–30.

22. NIH. [cited 2015 Sep 21] Search Query: "respondent driven sampling" OR "respondent driven sample" OR "respondent-driven sampling" OR "respondent-driven sample" [Internet]. Report. Database. 2015. Available from: http://tinyurl.com/m76bze2

23. White RG, Lansky A, Goel S, Wilson D, Hladik W, Hakim A, et al. Respondent driven sampling—where we are and where should we be going? Sex. Transm. Infect. 2012; 88:397–9. [PubMed: 23012492]

24. WHO. Introduction to HIV/AIDS and sexually transmitted infection surveillance: Module 4: Introduction to Respondent Driven Sampling. [Internet]. Geneva, Switzerland: World Health Organization; 2013. Available from: http://www.who.int/iris/handle/10665/116864 [cited 2015 Sep 28]

25. Tomas A, Gile KJ. The effect of differential recruitment, non-response and non-recruitment on estimators for respondent-driven sampling. Electron. J. Stat. 2011; 5:899–934.

26. Wejnert C. Social network analysis with respondent-driven sampling data: A study of racial integration on campus. Soc. Netw. 2010; 32:112–24.

27. Yamanis TJ, Merli MG, Neely WW, Tian FF, Moody J, Tu X, et al. An empirical analysis of the impact of recruitment patterns on RDS estimates among a socially ordered population of female sex workers in China. Sociol. Methods Res. 2013; 42:392–425.

28. Merli MG, Moody J, Smith J, Li J, Weir S, Chen X. Challenges to recruiting population representative samples of female sex workers in China using Respondent Driven Sampling. Soc. Sci. Med. 2015; 125:79–93. [PubMed: 24834869]

29. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annu. Rev. Sociol. 2001:415–44.

30. Crawford, FW., Aronow, PM., Zeng, L., Li, J. [cited 2016 Feb 9] Identification of homophily and preferential recruitment in respondent-driven sampling. ArXiv151105397 Stat [Internet]. 2015. Available from: http://arxiv.org/abs/1511.05397

31. Newman ME, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E. 2001; 64:026118.

32. Moore C, Newman MEJ. Epidemics and percolation in small-world networks. Phys. Rev. E. 2000; 61:5678–82.

33. Goel S, Salganik MJ. Assessing respondent-driven sampling. Proc. Natl. Acad. Sci. 2010; 107:6743–7. [PubMed: 20351258]

34. Gile KJ, Johnston LG, Salganik MJ. Diagnostics for respondent-driven sampling. J. R. Stat. Soc. Ser. A Stat. Soc. 2015; 178:241–69.

35. Centers for Disease Control and Prevention. [cited 2016 Apr 27] National HIV Behavioral Surveillance System in Injecting Drug Users – Round 3: Operations Manual [Internet]. 2012. Available from: http://www.cdc.gov/hiv/pdf/nhbs-idu3operationsmanual.pdf

36. UNAIDS. [cited 2017 May 5] Key Populations Atlas [Internet]. 2017. Available from: http://www.aidsinfoonline.org/kp/stock/data/CSV_DataFiles/AIDSinfo%20KP_2015_en.zip

37. Zhou T, Liu J-G, Bai W-J, Chen G, Wang B-H. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. Phys. Rev. E. 2006; 74:056109.

38. Kiss IZ, Green DM. Comment on "Properties of highly clustered networks". Phys. Rev. E. 2008; 78:048101.

39. Barabási A-L, Albert R. Emergence of scaling in random networks. science. 1999; 286:509–12. [PubMed: 10521342]

40. Khan B, Dombrowski K, Saad M, McLean K, Friedman S. Network Firewall Dynamics and the Subsaturation Stabilization of HIV. Discrete Dyn. Nat. Soc. 2013; 2013:e720818.

41. Ma X, Zhang Q, He X, Sun W, Yue H, Chen S, et al. Trends in prevalence of HIV, syphilis, hepatitis C, hepatitis B, and sexual risk behavior among men who have sex with men: results of 3 consecutive respondent-driven sampling surveys in Beijing, 2004 through 2006. JAIDS J. Acquir. Immune Defic. Syndr. 2007; 45:581–7. [PubMed: 17577125]

42. Valente TW. Network interventions. Science. 2012; 337:49–53. [PubMed: 22767921]
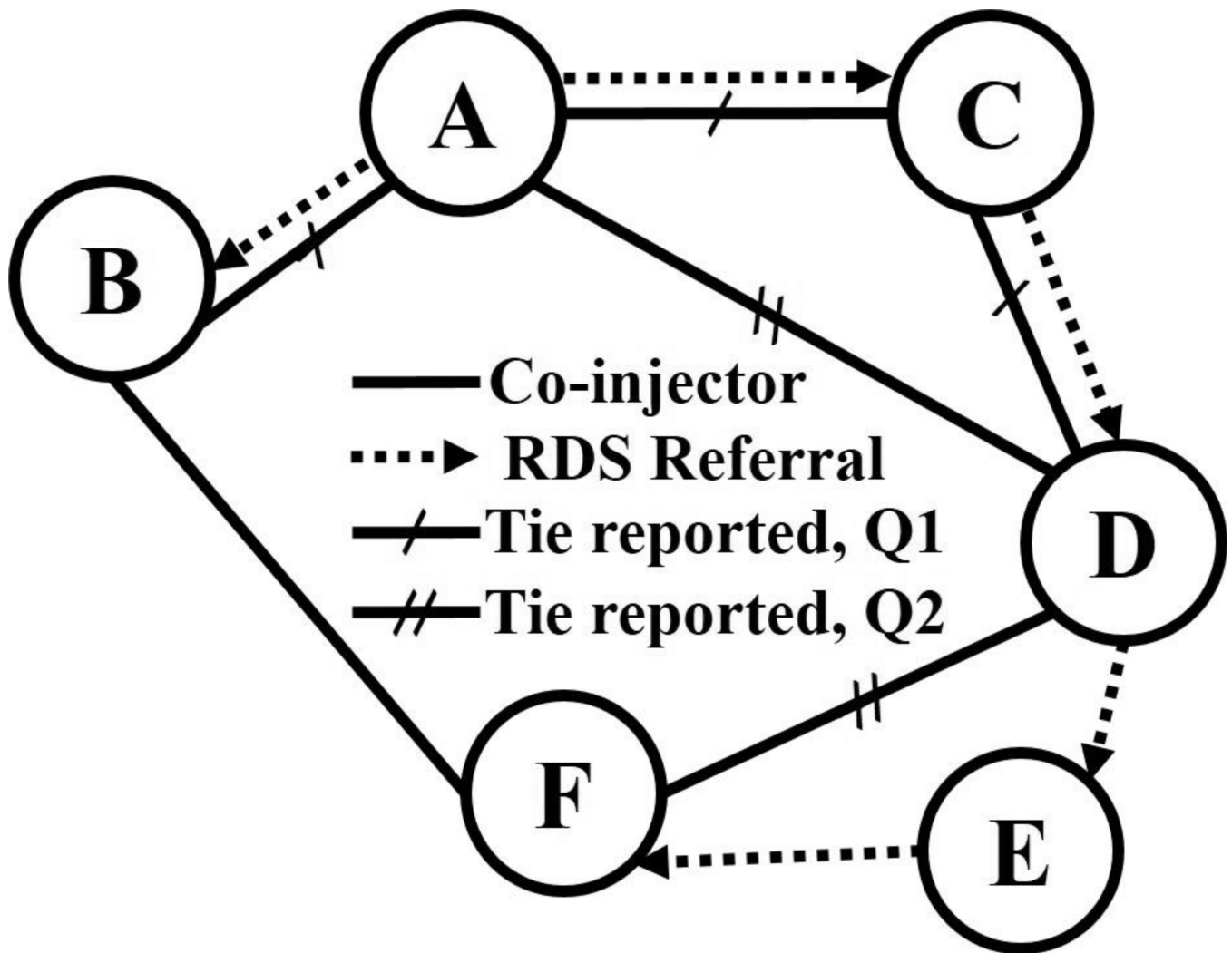
Figure 1.
A simplistic network to illustrate the clustering coefficient measure among co-injectors sampled with RDS. Person A co-injects with 3 people, making him involved in 3 co-injecting triplets (ABC, ABD, ACD). Since only one triplet (ACD) is closed for person A, he has a co-injecting network clustering level of 1/3. The co-injecting ties on the whole network shown have a local clustering coefficient of 0.278 and a global clustering coefficient of 0.333.

Table 1

Number of recruits in the RDS samples by wave and seed in each city. A non-recruiting seed in each city is omitted. Totals row exclude seeds.

| Wave | Cebu Seeds | | | | | | | | Mandaue Seeds | | | |
	9994	9995	9996	9997	9998	9999	8993	8994	8995	8997	8998	8999
0	1	1	1	1	1	1	1	1	1	1	1	1
1	2	2	2	2	2	2	2	2	2	2	2	2
2	4	4	4	4	4	4	4	2	2	4	2	2
3	3	5	6	8	8	8	3		4	5	2	4
4	4	6	7	10	10	16	2		4	4	3	4
5	1	4	8	12	18	30	2		6	1	5	4
6		2	5	9	16	42	4		3		6	4
7				8	19	48	7				3	8
8				4	16	28	6				3	12
9				2	10	24	5				3	12
10					4	7	4				2	16
11					1	4	4					18
12						1	7					21
13							6					15
14							4					9
15							6					8
16							8					4
17							6					1
18							2					
19							2					
20							2					
21							1					
Total	14	23	32	59	108	214	87	4	21	16	31	144

Table 2

Comparison of Demographic, Behavioral and Network Statistics in Cebu and Mandaue computed using the RDS Successive Sampling estimator [12]

	Cebu	Mandaue	Cebu-Mandaue
	Mean/%	Mean/%	Difference[a]
Friendship network degree	21	8	13***
HIV prevalence	52	34	18**
Age	32	30	2
Age (by category)			
15 to 24	28	42	−14*
25 to 34	32	27	5
35 and older	40	32	8
Educational Attainment			
No School	17	13	4
Elementary School	43	56	−13*
High School or Higher	40	31	9
Married	12	8	4
Years of injecting drug use	10	9	1
Age at first injecting drug use			
18 and under	34	36	−2
>18 to 25	40	46	−6
> 25	26	17	9
Injects at shooting gallery	83	80	3
Sharing with other PWID[b]	63	61	2
Did you inject with a used needle/syringe?[b]	59	64	−5
Did you pool funds with other PWID to purchase drugs?[b]	89	90	−1
Recruiter / Recruit Characteristics			
How long have you known your recruiter?			
< 1 year	14	20	−6
1–3 years	18	28	−10
> 3 years	69	53	16*
In the last 30 days, how often did you inject drugs with your recruiter?			
Did not inject with recruiter	23	20	3
Occasionally (less than 1× per week)	25	29	−4
At least 4 times (at least 1× per week)	52	51	1

Notes:

[a] two sample t-tests of the difference with

* indicating $p < 0.05$,

** indicating $p < 0.01$, and

***
indicating p<0.001.

b
Summary measures of injecting behaviors combine people who answered "Yes" to this question "at their last injection" or as "usual behavior in the past 6 months."

Table 3

Two measures of clustering coefficients in each city under different assumptions about missing data.

Missing data assumption	Local clustering coefficient		Global clustering coefficient	
	Cebu	Mandaue	Cebu	Mandaue
Triplet is open if missing	0.351	0.035	0.336	0.028
Triplet is closed if missing	0.377	0.036	0.346	0.037
Drop missing triplets	0.359	0.035	0.339	0.028
Average	0.363	0.035	0.340	0.031