

Full Length Research Paper

Collecting whole network data for human immunodeficiency virus prevention: A review of current strategies

Basmattee Boodram and Chyvette T. Williams

Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL, 60612, USA.

Accepted 11 March, 2013

The study of social networks is essential to understanding the spread of infectious diseases. This study reviews strategies for collecting whole (that is, sociometric) network data for human immunodeficiency virus (HIV) prevention. Using selected criteria, peer-reviewed journal articles published from 1980 to 2012 were searched in ISI Web of Knowledge and PsychInfo databases, and social network journals. Data from articles represent 12 whole network studies; the preponderance of the remaining articles was from personal (that is, egocentric) network studies. A common approach used to devise a whole network was recruiting and linking personal networks. Other approaches included venue-based linkages and use of a population roster. Ethnography and chain referral methods were key components of a multi-method approach to successfully acquiring a whole network. Few studies adequately explicate data collection and linking methodologies. Potential ways to augment and standardize reporting for similar studies are suggested.

Key words: Human immunodeficiency virus (HIV), social network, sociometric, egocentric, recruitment.

INTRODUCTION

The nature of human immunodeficiency virus (HIV) transmission will benefit from a clearer understanding of the role and dynamics of social networks in order to curb further spread of the disease. While social networks have been explored as a factor in the spread of HIV since the early years of the epidemic (Auerbach et al., 1984), growing evidence citing that individual risk behaviors alone do not explain racial disparities in HIV, other sexually transmitted infections (STIs) and blood-borne infections (BBIs) outcomes increases the need to understand network dynamics of infection (Adimora et al., 2009; Ellen et al., 1998; Hallfors et al., 2007). In a nationally representative study of young adults, racial disparities in the prevalence of STIs were not explained by substance use or risky sex

practices (Hallfors et al., 2007). Racial/ethnic minority groups are much more likely to belong to social and sexual networks with higher background STI/HIV prevalence, and this has been hypothesized to be a key contributing factor to HIV and STI disparities. With a growing interest in network research, combined with pressing need to prevent further spread of HIV, a paper that summarizes strategies for recruiting whole networks for HIV prevention is useful for (1) cataloguing the ways in which whole networks can be collected for HIV studies, (2) orienting new and emerging HIV network researchers, and (3) informing future research in this area.

In general, the majority of social network studies use either whole network (that is, sociometric) or personal

network (that is, egocentric) designs. Personal network studies involve focal individuals (“egos”) and those they are tied to (“alters”), while whole network designs attempt to understand social structure through a (not always) complete inter-connected web of direct and indirect ties among individuals in a bounded network. The manner in which research questions is conceptualized dictates the type of network studies conducted. This, combined with the incontrovertible difficulty and resource-intensiveness of conducting whole network studies, may account for the greater number of personal network studies in the literature. Moreover, numerous ethical issues are likely to arise in all network research, especially in HIV research which reports on sexual and drug use relationships (Woodhouse et al., 1995).

Studies examining social networks and HIV varied in the ways networks are incorporated into the study. Most commonly, network characteristics have been incorporated as data collection items on social factors contributing to disease incidence, prevalence, or behavior. This is typically done by characterizing risk networks with regard to structural, compositional, relational and functional aspects that influence HIV risk. For example, factors such as network size and attributes of both network members and relationships have been shown to influence risk behavior (De et al., 2008; Latkin et al., 2003; Smith et al., 2004; Williams and Latkin, 2007). Social network approaches have also been employed in the design of intervention studies that seek to alter behavioral norms within networks; one common method has been to use peer opinion leaders to diffuse new norms throughout the network (Latkin et al., 2003; Kelly et al., 2006; Latkin, 1999). Finally, social network approaches, including respondent-driven sampling (RDS), are also regularly used in the recruitment of hidden and/or high-risk groups (Abramovitz et al., 2009; Heckathorn, 1997; Heckathorn et al., 2002; Kendall et al., 2008).

Numerous articles have been published examining the effects of personal network characteristics on risk behaviors and HIV-related outcomes; and the field has benefited significantly from these contributions. Personal network studies, which predominantly use self-report data by the index about their direct network members, have largely been employed to understand interpersonal influence and risk factors and diffusion of norms from a peer leader to network members. However, personal network data are not optimally designed to address questions of population transmission and spread, unless the multiple, smaller networks they yield can be connected to form larger network components. Components represent the number of persons who can be reached through direct and indirect linkages in a network. Whole networks, by definition, are complete networks depicting all connections within a defined social group or geographic space; therefore, they are designed to yield larger components than those sought in personal networks. For this reason, they may be more informative when considering population-

level inter-relationships, risk connections and potential infection spread. Because they yield larger interconnected structures, whole networks offer multiple viewpoints for identifying prevention points, for example, through measures of centrality, position, exposure and influence.

Simulation studies, such as those conducted by Morris et al. (2009) have used population parameters from large representative studies to illustrate the dynamics of infection spread under conditions where concurrent partnerships exist. These types of studies are aptly suited for demonstrating concurrency and visual representation of where risk is located within a social network. Morris et al. (2009) showed that even modest levels of concurrency greatly impacts epidemic potential and also demonstrated how sexual mixing patterns by race could have produced present-day racial health disparities. Although simulation studies are informative in explaining observed patterns and to predict future patterns, they do not negate the need to understand networked relations only possible with empirical field studies. Being able to identify and prevent new HIV infections, when persons are most infectious, is critical to reducing prevalence.

Understanding network dynamics in empirical field studies is hindered by multiple challenges to collecting whole network data for HIV prevention. First, recruiting whole networks depends upon having clearly delineated boundaries for sample recruitment. For studies within organizations (e.g. schools), it is possible to identify all members from a roster or similar mechanism and proceed to determine ties among them. For population-based recruitment in high prevalence communities, social and geographic boundaries are not always clear. For example, persons living with or at risk for HIV may occupy multiple geographic spaces (e.g. living in an area distinct from the one where they engage in risky behaviors). Alternatively, the target population may be socio-demographically diverse and include members from both hidden and unhidden populations. Many HIV studies target individuals based on their risk behaviors or membership in a defined high-risk group. In contrast, whole network studies are comprised directly and indirectly linked individuals who may be at varying levels of risk for infection. Under a network perspective, persons generally considered “low risk” and who may not otherwise be targeted for prevention or intervention remain at risk of infection through indirect relationships. The combination of one’s position within a network and the behaviors of direct and indirect ties may affect one’s behaviors and health outcomes.

Enumeration and recruitment of network members represent two additional challenges to sampling whole networks. Enumeration can be hindered by multiple biases, including recall and misclassification biases, masking (that is, limiting disclosure as a means of protecting privacy) and inability to identify anonymous partners. Recruitment of alters can be potentially more challenging than enumeration, particularly among connected members

who are at risk of HIV through indirect ties. Finally, depending upon the recruitment design, identifying linkages among individuals will be necessary to complete the network, and this can be done through several methods, including direct observation and respondents reporting the ties among their network members. Achieving a complete network for HIV prevention is methodologically challenging and very rare. For most cases in HIV research, a whole network refers to a network that recruits beyond alters who are connected directly or through one intermediary person to include longer indirect linkages and the connections among them.

Multiple strategies have been used to recruit whole networks for HIV research with prevention implications, but no systematic approach currently exists. Often, a mixed method approach that combines quantitative and qualitative methods is required to recruit a whole HIV risk network. Specifically, we address the following questions: What are the strategies used to sample, recruit, and collect whole network data for HIV prevention? What are the strengths and limitations of the strategies used? Are there recommendations that can be made toward standardizing approaches for specific goals and/or populations?

METHODOLOGY

Inclusion criteria

Studies were selected for review if they were: quantitative STI or HIV research that collected whole network data and were published in a peer-reviewed journal between January 1, 1980 and March 30, 2012. Since achieving whole networks in the truest sense in HIV studies is very rare because of the methodological challenges indicated, we used a more encompassing criterion. Granovetter (1973) suggests that personal networks are made up of an ego, ego's contacts and their contacts. As such, in this study, we report on studies whose recruitment went beyond these two levels of ties. Our interest was specifically in social network studies, rather than studies employing primarily cluster analyses, contact-tracing, respondent-driven sampling methods, phylogenetic studies, and investigations of outbreaks; as such, studies focused on these latter methods of recruitment, analysis, or linking were excluded.

Our search was limited for two reasons. First, our focus is on understanding field-based whole network studies. While the aforementioned methods may share some properties of networks, they represent different approaches that may not be as applicable to field-based research. For example, investigators seeking to recruit whole networks within high-risk communities usually do not do contact-tracing or investigate outbreaks; these typically have been functions relegated to public health departments. Moreover, molecular phylogenetic studies seek to discover the source of infection, and depend on having a recruited sample to trace molecular data. Furthermore, cluster analysis assigns a set of individuals to a group or cluster based on some algorithm (e.g. distance). Finally, respondent-driven sampling is the method that yields dendritic (that is tree-like) structures upon which limited network analyses can be performed. Second, given the large number of studies of HIV, STI, and other blood-borne infections that now incorporate a network component, we necessarily had to create boundaries for this review, and we acknowledge that these boundaries exclude many network studies. Other eligibility criteria

were: a focus on either drug or sexual networks and field-based empirical studies. Conceptual, simulation and internet-based studies were excluded. Internet-based studies were excluded because they are categorically different from field-based studies and have their own specific set of methodological issues. Studies conducted or published outside of the United States were included in the search; however, non-English language publications were not included in the research.

Search strategy

Two databases (ISI Web of Knowledge and PsychInfo), as well as 3 social network journals (Connections, Social Networks, and the Journal of Social Structure) were searched for relevant articles. ISI Web of Knowledge is the largest database of citations in the Arts and Humanities, Social Sciences and Sciences. Both the Social Science Index and Medline are indexed within this database. Subject searches were conducted using the terms "HIV" or "STI" and "network." Additional searches were conducted using the terms "network" and "gonorrhoea" or "Chlamydia" or "syphilis" or "sexually transmitted" and "network" and "hepatitis." The search terms were purposefully chosen to be broad in hopes to capture all indexed studies on HIV, STIs, blood-borne infections and networks.

Analysis

The Web of Knowledge searches yielded a total of 2,248 articles for the time period (including duplicates between the two searches), of which 65 were selected for more in-depth screening for inclusion. The PsychInfo search yielded 1,050 articles, of which 30 were selected for further review. The journals Social Networks, Journal of Social Structure, and Connections yielded 129 articles that were reviewed; of which 14 were selected for further review. Article titles, and/or abstracts were reviewed to generate list of articles for further review and final list of potential articles was retrieved, reviewed and agreed upon by both authors. Of note, our results are an attempt to report on distinct studies, not specific analyses or articles written within each study. Figure 1 summarizes the article search flow.

RESULTS

The final number of whole network studies meeting the inclusion criteria was 12 (Figure 1). The geographic variation, population studied, sampling, recruitment and data collection methods varied across the 12 studies (Table 1), as well as key features such as largest network component, number of contacts recruited, and linkage procedures (Table 2). This study describes some of these features in more detail subsequently.

Regional spread, risk groups and outcomes targeted among identified studies

Several studies were conducted in multiple regions. Of the 12 studies identified, 5 were conducted outside of the United States, 1 in Africa (Malawi), 1 in Asia (Hong Kong), 2 in Canada (British Columbia, Manitoba), 1 in Puerto Rico, and 1 in Europe (Russia, Bulgaria, Hungary). Of the studies conducted within the United States, 2 were conducted in South-eastern states (Atlanta,

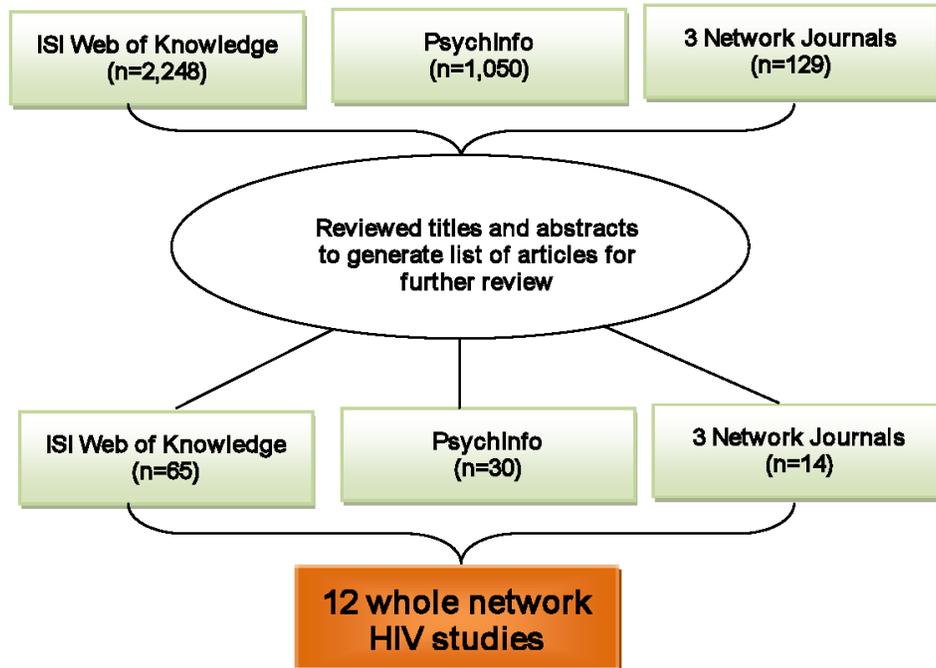


Figure 1. Literature search flow^a.

GA), 2 in southwestern states (Colorado Springs, CO, Flagstaff, AZ), 1 in the west (Los Angeles, CA), and 3 in the North-eastern region (Brooklyn, NY (2); Hartford, CT). Five studies targeted men who have sex with men (MSM), although perhaps not exclusively; 1 targeted female sex workers; and 6 targeted drug users. All of the studies examined both sex and drug risks regardless of primary target population. Additionally, all of the studies had HIV outcomes as the primary interest, though STIs were reported on. No whole network studies of hepatitis infection were found, but studies targeting injection drug users (IDUs) did report on hepatitis prevalence and risk.

Sampling and recruitment of whole HIV networks

Table 1 summarizes the key components of the 12 whole network studies identified along with the main references that were identified in our searches (Amirkhanian et al., 2009; Friedman et al., 1997, 1999, 2007a, b, 2008; Helleringer and Kohler, 2007; Helleringer et al., 2009; Klovdahl et al., 1994; Lee et al., 2009a, b; Liebow et al., 1995; McGrady et al., 1995; Remple et al., 2007; Rice et al., 2012; Rothenberg et al., 2001, 2000; Weeks et al., 2002; Wylie et al., 2007). Sampling and recruitment of whole networks was rarely achieved by a priori delineation of social or geographic boundaries. Although, there was variation among how the studies recruited and constructed whole networks, we identified 4 main strategies or designs: (1) personal networks (7 studies), (2) random walk (2 studies), (3) census (1 study) and (4)

two-mode (2 studies). We review each strategy in more detail subsequently.

Recruiting personal networks via targeted sampling

Tables 1 and 2 list and provide details on the studies that used personal networks to build whole networks. Devising a whole network by recruiting and linking personal networks was by far the most common approach. In most cases, recruitment began with outreach in targeted locations where the population(s) of interest could be sampled. Three of the studies began recruitment in specific venues: the study of indoor female sex workers (FSW) in British Columbia began recruitment in sex establishments; the study of runaway homeless youth (RHY) recruited from a drop-in center for homeless youth using an event-based approach; and the Central European studies of MSM began recruitment in bars and clubs. Venues offer clear-cut boundaries for beginning recruitment, but may bias the sample toward individuals who frequent those venues. Outreach and recruitment in community areas with high rates of HIV risk behaviors is another common approach and may allow researchers to intercept potential participants in their “natural” environment. Venue-based sampling, on the other hand, is useful when there is no particular community location where members of a high-risk group congregate.

For many studies, outreach, ethnography and chain referral methods were key components of a multi-method approach to successfully acquiring a whole network sample.

Table 1. Description of whole network studies, by recruitment strategy.

Study, time frame, geographic location	Study population/risk group	Sampling, recruitment and data collection methods
Personal networks linked to form a whole network		
Colorado Springs Study (1988-1992; Colorado Springs, CO, USA)	High risk heterosexuals (that is, IDUs, prostitute women); 1 st year: (n=255 index) and their sexual injection networks (n=1,296)	Participants were recruited via clinics, drug treatment centers, testing sites, and through outreach. Recruited participants (egos) were interviewed to nominate social, sexual, illicit drug and injecting drug partners for the prior 6 months. Personal networks linked using health department charts.
High Risk Settings Study (1998-1999; Hartford, CT, USA)	Adult drug users (n=293 index); mean age=37.3 years	Ethnography to document hidden drug use sites. Street outreach in high drug use neighborhoods-targeted sampling. Respondent-driven methods to ascertain personal networks. Personal networks linked using photos to ascertain cross-network memberships.
Indoor Female Sex Workers Study (June 2004- February 2006; Vancouver, British Columbia, Canada)	Female sex workers (n=49) and their recent sexual partners (n=504); mean age=37 years	Targeted sampling to select ethnically and geographically diverse sex establishments (n=7) from a list that was compiled using outreach methods. Female sex workers at each establishment were asked to identify and complete a proxy questionnaire for up to 10 recent sexual partners. Personal networks were linked to create a whole network through clients who patronized ≥ 1 commercial sex establishments (house regulars)
Social Factors and HIV Risk Study (July 1991-January 1993; Brooklyn, NY, USA)	IDUs and their drugs and sex partners (n=767 index)	Street outreach and ethnography in areas of heavy drug use in a defined area. Chain referral to ascertain personal networks. Personal networks were linked using names provided. Linkages confirmed between two individuals if 1) engaged in face-to face contact with research staff at the same time, 2) were observed together in public.
Young MSM Networks in Central Europe (2003-2004; St. Petersburg, Russia; Sofia, Bulgaria; and Budapest, Hungary)	Young MSM (n=156); age=varied by city/analytic sample	Ethnography used to select venues (bars, night clubs), index participants, and to observe social circles and leaders within circles. Initial "seed" recruited and interviewed to identify and recruit all members of immediate friendship group. Network leaders were identified for intervention.

Table 1. Contd.

Networks, Norms, and HIV Risk Among Youth (2001-2004; Brooklyn, NY, USA)	Injection drug users and young adults were index cases (n=112) and network members (n= 353); age 18 to 24; of all participants, 71% Latino, 20% African-American; 15% MSM, 43% injected drugs	Ethnography, focus groups, and survey data used. Three waves of recruitment (index, partners of index, and network members of partners). IDUs and young adults recruited separately. Young adults recruited using a population-based representative sample from a previous study and targeting of specific households. IDUs (non-representative) targeted specific venues (e.g. shooting galleries, drug purchasing venues, needle exchange sites) as well as relied on walk ins. Participants were asked to bring in network members for interview, or given them a coupon to be redeemed, or were located by project staff.
Runaway Homeless Youth (2008; Los Angeles, CA, USA)	Young individuals age 13 to 24 (n=136) seeking services at an agency serving homeless adolescents	Direct recruitment of 150 participants seeking services at a specific agency during the recruitment, of which 136 (90.7%) participated. Self-administered computerized interviews to collect individual-level data on index and a face-to-face interview to generate list of, and data on, network members.
Random walk		
Atlanta and Flagstaff Networks Study (1995-1999; Atlanta, GA and Flagstaff, AZ, USA)	Individuals at risk for HIV due to drug-using or sexual practices, with emphasis on interaction of risks (n=228 index); mean age=40.3	Ethnography used to ascertain community sites. Two seed participants from three geographically separated community sites selected to start chain. Chain-link referral and random walk methods to ascertain community chains of ~10 contacts.
Atlanta and Puerto Rico Ethnic Youth Network (Study period unreported; Atlanta, GA, USA and San Juan, Puerto Rico)	Adolescents 12-15 residing in defined geographic 1990 Census tract areas selected to define an ethnically homogenous, lower SES status region (n=43 index in Atlanta, n=52 index in San Juan). Atlanta: 129 respondents, 1141 total nodes, 12 components	Simple random sampling in defined areas using 1990 census data. In Atlanta, a roster of public of public housing project residents along with a two-stage method was used identify index. In San Juan, a random selection from a door-to-door canvas in a public housing project used to identify index. Respondents generated list of respondents (variable length) using 8 name generators. To continue a random walk, the next candidate for interview was randomly selected based on eligibility criteria (age 12-29 and resided in metropolitan area).
Census		
Likoma Network Study (October 2005-March 2006; Likoma, Malawi)	Likoma residents (n=923); age=18 to 35 years	Census of entire island to obtain roster of all households. Sexual network and health survey among study population. Biomarkers of HIV infection collected among network survey participants. Whole sexual network on island linked all individuals.

Table 1. Contd.

Two-mode networks		
Winnepeg Social Network Injection Drug Use Study (December 2003-August 2004; Winnepeg, Manitoba, Canada)	IDUs (node 1) (n=435 index); aged 15-64 (median=36 [male] and 35 [female]); 56% male; 47% First Nations or North American Indian, 34% White, 15% mixed race/ethnicity; Hotels (node 2): Up to 6 specific hotels where individuals inject drugs was identified by n=172 individuals	Advertisement and word-of-mouth in IDU neighborhoods. Individuals were interviewed using a structured questionnaire. Participants identified specific hotels (typically single-room occupancy or residential hotels) where they inject drugs. Two mode network linked up individuals through hotels identified.
Integrated Treatment Center Clinic (July 2007-March 2008; Hong Kong)	Chinese, MSM, recently HIV-infected (<1 year) clinic patients (n=114 recruited); mean age=33.5-37.5 years (based on papers published on sub-samples)	Two-mode affiliation networks converted to 1-mode networks. Self-administered questionnaire inquired about partners and locations where index clients frequented for seeking sex partnership (partner sourcing). Venue-based linking-individuals linked to partner sourcing venue, and then individuals linked to each other. HIV-1 pol gene phylogenetic analysis used to augment linkage.
Likoma Network Study (October 2005-March 2006; Likoma, Malawi)	Likoma residents (n=923); age=18-35 years	Census of entire island to obtain roster of all households. Sexual network and health survey among study population. Biomarkers of HIV infection collected among network survey participants. Whole sexual network on island linked all individuals.
Winnepeg Social Network Injection Drug Use Study (December 2003-August 2004; Winnepeg, Manitoba, Canada)	IDUs (node 1) (n=435 index); aged 15-64 (median=36 [male] and 35 [female]); 56% male; 47% First Nations or North American Indian, 34% White, 15% mixed race/ethnicity; Hotels (node 2): Up to 6 specific hotels where individuals inject drugs was identified by n=172 individuals	Advertisement and word-of-mouth in IDU neighborhoods. Individuals were interviewed using a structured questionnaire. Participants identified specific hotels (typically single-room occupancy or residential hotels) where they inject drugs. Two mode network linked up individuals through hotels identified
Integrated Treatment Center Clinic (July 2007-March 2008; Hong Kong)	Chinese, MSM, recently HIV-infected (<1 year) clinic patients (n=114 recruited); mean age=33.5-37.5 years (based on papers published on sub-samples)	Two-mode affiliation networks converted to 1-mode networks. Self-administered questionnaire inquired about partners and locations where index clients frequented for seeking sex partnership (partner sourcing). Venue-based linking-individuals linked to partner sourcing venue, and then individuals linked to each other. HIV-1 pol gene phylogenetic analysis used to augment linkage.

SES, socioeconomic status; IDU, injection drug user; MSM, men who have sex with men.

Table 2. Key features and characteristics of 12 whole network studies, by recruitment strategy.

Study, time frame, geographic location	Largest network component ^a	No. of contacts respondent could recruit or report on	Contacts recruited or proxy reports provided by respondent	Response rate for recruited contacts	Link identification
Personal networks linked to form a whole network					
Colorado Springs Study (1988-1992; Colorado Springs, CO, USA)	Over 600	All close personal sexual and injection contacts within past 6 months	Proxy reports and data pooling	75%	Names, demographic and locating information for each network member; matches were based on name, alias, ethnicity, gender, approximate age, and agency attendance; respondent asked about known relationships between each pair of associates using a matrix format
High Risk Settings Study (1998-1999; Hartford, CT, USA)	193 (66% of sample)	2 drug-using network members recruited	Contacts recruited	Not found	Full names and/or nicknames, pictures
Indoor Female Sex Workers Study (June 2004- February 2006; Vancouver, British Columbia, Canada)	553	Up to 10 most recent sex partners	Proxy reports	Not applicable	Mentions of same partners that match on race, residence, age, physical description
Social Factors and HIV Risk Study (July 1991-January 1993; Brooklyn, NY, USA)	277 (for sample of drug users); "large" (for group sex analysis)	Up to 10 risk network members	Mixed	Not found	First or street name, "questionnaire and other data"
Young MSM Networks (2003-2004; St. Petersburg, Russia; Sofia, Bulgaria; and Budapest, Hungary)	65	All friends identified and recruited	Contacts recruited	78.7% participation rate	Peer-driven recruitment
Networks, Norms, and HIV Risk Among Youth (2001-2004; Brooklyn, NY, USA)	206	Named network members up to a limit of 10 sex partners, 5 IDU partners, and two group sex event attendance partners	Contacts recruited	Recruited until sample goal reached (e.g. n=200 for IDUs)	Participants considered link if one named the other as someone whom, during the prior 3 months, they had a) injected drugs, b) had sex with and/or, c) attended a group sex event with.
Runaway Homeless Youth (2008; Los Angeles, CA, USA)	K-core range: 0-7 Large number of isolates or those with only 1-tie (56.6%)	All people interact with: at agency, hang out with, have sex with it, etc	Proxy reports	Not applicable	First and last initial, aliases, age, gender, race/ethnicity and attendance at agency where recruitment occurred for index Event-based approach (EBA): Regulars of a socio-physical space linked up

Table 1. Contd.

Random walk					
Atlanta and Flagstaff Networks Study (1995-1999; Atlanta, GA and Flagstaff, AZ, USA)	228 were members of 6 community chains; 17-94% of group members	All social, sexual, drug-using, and needle-sharing	Contacts recruited	40% of contacts interviewed	Linked via recruitment Names, general locations, descriptions used to locate associates. Respondent asked about relationships among pairs of contacts
Atlanta and Puerto Rico Ethnic Youth Network (Study period unreported; Atlanta, GA, USA and San Juan, Puerto Rico ^b)	Atlanta: 763	Respondents generated variable list of network members interviewed	Contacts randomly selected for interview	Atlanta: Of 43 walks, 38 (89%) completed 3-step, 2 (5%) completed 2-step walks, and 1 (2%) completed a 1-step	Linked via recruitment Names, addresses, phone numbers used to locate associates
Census					
Likoma Network Study (October 2005-March 2006; Likoma, Malawi)	883	Up to 5 sex partners	Sex partners identified via roster and recruited	Not applicable (88% for respondent)	First names or initials, plus demographics; migration rosters
Two-mode networks					
Winnipeg Social Network Injection Drug Use Study (December 2003-August 2004; Winnipeg, Manitoba, Canada)	Component consisted of 154 IDUs and 39 hotels; K-core max size=3 (range 1-3)	Sample of 435 IDUs identified up to 6 specific hotels; 49 distinct Winnipeg hotels included in analysis (from 162 Winnipeg residents)	Individuals interviewed about specific hotels	Not reported	Individuals were linked to hotels; only indirectly to each other
Integrated Treatment Center Clinic (July 2007-March 2008; Hong Kong)	19 of 91 persons were not connected	All sex partners in 3 years prior to infection	Proxy data on locations where partners sought	Not applicable	2-mode matrix of MSM and locations converted to 1-mode matrix linking MSM

^aComponent size is the number of persons who are linked through direct and indirect linkages. ^bIn two other locales, Seattle, Washington and Honolulu, Hawaii, a more limited scope of related research focused on adolescents' social relations. Data not reported here. HIV, human immunodeficiency virus; MSM, men who have sex with men.

sample. The Social Factors and HIV Risk study (SFHRS) and High Risk Settings study are good examples of this multi-method approach. Table 1 indicates which studies conducted ethnography as

part of their recruitment approach. Outreach and ethnography in high-risk areas are labor-intensive, but elucidate the culture and context of the target groups and develop rapport, trust, and credibility.

These high-risk areas become the source of initial recruits who become the "index" persons in the net-work study. Index persons are then asked to name their network members, who may or may not

all be recruited.

The Central European studies of MSM conducted targeted sampling and devised a method to identify the most central person in a small “social circle.” Another approach has been to simply assign the initial recruits to “index” status. The former approach of determining the index via nominations from members of a social circle may be more important for interventions than for more descriptive epidemiologic studies of whole networks. The landmark Colorado Springs study and the Networks, Norms, and HIV Risk among Youth (NNAHRAY) recruited multiple risk groups and conducted detailed network inventories to ultimately reveal linkages across distinct subgroups.

Random walk approach

The Atlanta and Flagstaff Networks and the Atlanta and Puerto Rico (PR) Ethnic Youth studies targeting ethnic youth employed the random walk design (Tables 1 and 2). Random walk designs are link-tracing designs where each successive addition to the chain-link identified by randomly selecting an associate from the list of current recruits and locating them for study inclusion. In the Atlanta/PR study, a two-stage probability sampling of adolescents was conducted in census tracts, which was then followed by simple random sampling of named associates. With this design, members of the sample are linked via their recruitment. Therefore, unlike the personal network design, no linkages need to be made after subject recruitment. The Atlanta/PR study asked respondents to provide the name, address, and telephone number of associates, whereas the Atlanta and Flagstaff studies requested “general” location data. The Atlanta/PR study was conducted in the early 1990s; requesting name, address, and telephone number might present challenges with institutional review boards for investigators seeking to use this approach today. A limitation with the random walk design is the difficulty in locating associates with the locator information obtained from respondents.

Census approach

The Likoma Network study (Malawi) represents the rare opportunity to capture a complete whole network with the availability of a roster of the entire island’s population as well as the use of biomarkers of HIV infection to validate linkages (Tables 1 and 2). The Likoma study produced the largest component among the studies; but it represents an exception since it used a census approach in a bounded geographic area (Likoma Island) (Helleringer and Kohler, 2007; Helleringer et al., 2009). Most HIV prevention studies looking to recruit whole networks are conducted under less optimal circumstances, where geographic boundaries are ambiguous and a population census

is difficult to acquire.

Two-mode network approach

Another approach represented in the studies identified was venue-based linking (also known as two-mode networks), whereby individuals were linked through attendance at common venues. The Winnipeg Social Network Injection Drug Use study (SNS II) and Integrated Treatment Center Clinic study (Hong Kong) employed this strategy. The SNS II collected personal network data and did not create linkages among the IDUs in the study. Instead, a whole network was constructed by linking each IDU to a hotel where they injected. IDUs were only indirectly linked via their joint “membership” at the same hotel. The Integrated Treatment Center Clinic study conducted a similar analysis, but created “direct” linkages among the MSM in the study by assuming linkages among men who reported attending the same venue. The SNS II study has the analytic capability to do this as well. A limitation of these studies is that direct connections among individuals are assumed based on joint attendance at a venue or event. This approach might be useful when person-to-person linkages cannot be collected.

Data collection and link identification

Table 2 summarizes details on several key features of the 12 network studies, including how network contacts were recruited, the largest component size and link identification. Component size is the number of persons who are linked through direct and indirect linkages and is important in determining the degree to which infection can spread throughout a population. Many of the studies allowed reporting of all contacts, but may have recruited only a subset. Recruiting all contacts is very resource-intensive; and therefore, many studies may limit the number of recruits while simultaneously collecting proxy data on all contacts. Proxy data allows a way to collect information on associates from index persons and pool data on the attributes of associates to determine links. Large components can result from this method, as evidenced by the Colorado Springs study and the FSW study. Names (first, street, or initials) and other identifying attributes were consistently used as a key descriptor for pooling and making linkages. Response rates for studies varied somewhat. Having to locate associates may severely affect the ability to recruit associates into the study. Using peer-driven recruitment approaches can increase enrollment of network associates into studies.

DISCUSSION

Since the mid 1980s, social network concepts have been applied to HIV research and social network methodologies

are increasingly being utilized to understand the HIV/AIDS epidemic. The justification for using a network approach in HIV research and in developing interventions resides in the social context in which individuals engage in HIV risk behaviors and in which they become exposed to HIV (Rothenberg et al., 1996).

More detail needed in HIV network studies

Our review suggests a need for more detail and specificity in titles and descriptions for social network studies, particularly given the growth in the number of network studies over the past two decades. Social network methods are particularly complex and a wide range of questions and analytic techniques can be generally classified as network-oriented. One of the primary challenges in classifying studies for this review was the lack of specific information on the type of network study being conducted. Specifically, explicit classification as to whether studies, and particular analyses, focused on personal networks or whole networks or both was rare. In some cases, the combined information provided by multiple articles on the same study was necessary to make this determination. While the term social networks or social network analysis is the umbrella term within which all network-oriented investigations fall, there are important distinctions between the goals and types of studies. In our review, the preponderance of studies identified evaluated solely personal networks, including studies of social support and social influence.

Perhaps more important than making an explicit statement regarding the type of study was providing specific details on sampling and recruitment, particularly for community-based whole network HIV studies where boundaries and sampling frames are often ambiguous. Significant variation existed in the explication of the sample recruitment method employed. How were boundaries determined, if at all? For example, a common practice was to use a personal network sampling design and link members to create a whole network. In this case, geographic boundaries may not need to be determined a priori. The question then becomes: how was the network initiated, with whom, and why?

Details on data collection and linkage construction, which is the process of determining the connections among network members and thus, creating the whole network, were also rarely comprehensive. What types of name generators were used and were there limits and a rationale for the number of alters individuals could report? What data were collected in order to determine linkages? Where did the network end and why? Academic journals limit the extent to which details of network studies, which tend to be complex and jargony, can be presented; and this may account for the consistent lack of detail. In some cases, articles were also published in books where considerably more detail was provided. Growing interest

in the use of network analysis could benefit from a clearer understanding of the design and methods of network studies. Investigators of network studies might consider publishing a methods paper that describes the study and which can be used as a reference for researchers wishing to duplicate or build on such methods. This will be especially useful for new network investigators seeking to replicate and advance network-oriented HIV studies.

Toward developing standards

Results of the search strategy suggest some ways that whole network studies might be approached based on the population of interest. Whether the target population represents a hidden or accessible group may offer a natural delineation in methodological approach. For example, risk groups such as MSM and IDUs are hidden populations, because of their stigmatized and illegal behavior, respectively. Their marginal status tends to make their networks more cohesive, either geographically or socially. Targeted sampling appears to be standard practice as a first step in locating hidden populations. Because IDUs tend to be low-income, they may live in poorer areas, while MSM may congregate at gay venues or gay-friendly spaces. Ethnography is then conducted within target areas to understand social dynamics among members. Ethnography is time-consuming, but builds trust and rapport that is important in garnering full participation of the target group and in validating linkages (Friedman et al., 2008, 1997; Weeks et al., 2002). From targeted areas, individuals can be recruited and their network members named and then linked. Additionally, it seems important to begin networks with individuals with high degree (that is, relatively large number of direct ties) to increase the chance of constructing sizeable whole networks. Proxy reports also appear to yield networks with large components. The Colorado Springs study demonstrated the possibilities of recruiting a large connected component from distinct but overlapping risk groups such as IDUs and commercial sex workers (Klovdahl et al., 1994).

Determining network ties among diverse and dispersed populations, such as high-risk heterosexuals, will likely require a different approach. High-risk heterosexuals constitute a large population of different genders and ages. They are defined by Centers for Disease Control (CDC) as anyone with a known risk (e.g. HIV infection, drug injection, male-to-male sex) in a heterosexual partner. Other studies focused on high-risk heterosexuals may define the category according to specific individual-level behavioral risk. For example, research on age mixing indicates elevated risk of transmission between young women and older men (Kraut-Becher and Aral, 2006). High-risk heterosexuals (e.g. women) may engage in relatively lower levels of risk, but be networked through sex

partners to groups with higher prevalent or greater risk of HIV infection than themselves. Therefore, a viable approach may be targeting groups that link with high-risk heterosexuals, such as behaviorally bisexual men and persons with a history of incarceration or drug use, and identifying their sex and drug partners and any linkages among them. The Social Factors and HIV Risk study, Colorado Springs study, Atlanta and Flagstaff Network studies exemplify this strategy with their attempts to recruit diverse samples.

On using chain-link, contact tracing for network purposes

Recruitment methods that generate chain links, e.g. RDS, are often easier to implement, but do not represent “true” whole network data. While chain-link samples are networked data in the sense that all individuals are connected to at least one other person in the sample, they are primarily recruiting designs with imposed structures. Further, they typically do not include triads or cross-links (that is, respondents linked through a commonly reported associate). Instead, these methods result in long, linked chains of individuals in a dendritic structure. Because the structure of the network is imposed by the recruitment design, certain network analytic techniques cannot be performed. An overview of the types of social network questions that can be addressed and analyses performed with RDS data has been recently reviewed (Wejnert, 2009). However, a possible, but currently underutilized option is to incorporate network inventories into data collection for RDS studies. This would augment the data by capturing information on networks that might not be captured in the linked sample.

Contact-tracing studies are another type of network-oriented design identified in the search. Typically, these studies fall within the purview of local health departments and employ network methods for investigating sources and spread of disease in an outbreak. While these studies use network methods, they are a form of post-hoc analysis attempting to link incident cases to the sources of spread. While many investigators planning network studies do not begin with contact-tracing data, such data have been used to understand HIV risk networks and to inform prevention strategies.

STUDY LIMITATIONS

This review has several limitations. First, our inclusion criteria were very specific and primarily limited to field-based empirical whole network studies originating in community settings. Second, our search terms may not have retrieved all whole network studies fitting the criteria due to variable keyword indexing. Furthermore, difficulties encountered in classifying papers due to lack of sufficient details may have led to the omission of relevant

research studies.

CONCLUSIONS

Whole network studies can provide unique opportunities to understand and interrupt the flow of infectious diseases in communities. This review suggests that no single recruitment strategy is sufficient for recruitment of a complete network. Each of the 12 studies reviewed had notable strengths as well as weakness. To optimize success of a near complete whole network, multiple methods are needed and these methods should be clearly delineated. More studies that include control or comparison networks are also needed to increase our understanding in differences in networks between different sub-populations. Understanding the structure and dynamics of social and risk networks remains critical for developing effective intervention to prevent and interrupt spread of infectious diseases.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Carl Latkin for his thoughtful review of the manuscript.

ABBREVIATIONS

HIV, human immunodeficiency virus; **IDUs**, injection drug users; **MSM**, men who have sex with men; **STI**, sexually-transmitted infections; **BBI**, blood-borne infections.

REFERENCES

- Abramovitz D, Volz EM, Strathdee SA, Patterson TL, Vera A, Frost SD, Proyecto E (2009). Using respondent-driven sampling in a hidden population at risk of HIV infection: who do HIV-positive recruiters recruit? *Sex Transm Dis.* 36(12):750-6.
- Adimora AA, Schoenbach VJ, Floris-Moore MA (2009). Ending the epidemic of heterosexual HIV transmission among African Americans. *Am. J. Prev. Med.* 37(5):468-71.
- Amirkhanian YA, Kelly JA, Takacs J, Kuznetsova AV, DiFranceisco WJ, Mocsonaki L, McAuliffe TL, Khoursine RA, Toth TP (2009). HIV/STD prevalence, risk behavior, and substance use patterns and predictors in Russian and Hungarian sociocentric social networks of men who have sex with men. *AIDS Edu. Prev.* 21(3):266-79.
- Auerbach DM, Darrow WW, Jaffe HW, Curran JW (1984). Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact. *Am. J. Med.* 76(3):487-92.
- De P, Cox J, Boivin JF, Platt RW, Jolly AM (2008). Social network-related risk factors for bloodborne virus infections among injection drug users receiving syringes through secondary exchange. *J. Urban Health* 85(1):77-89.
- Ellen JM, Aral SO, Madger LS (1998). Do differences in sexual behaviors account for the racial/ethnic differences in adolescents' self-reported history of a sexually transmitted disease? *Sex Transm. Dis.* 25(3):125-9.
- Friedman SR, Bolyard M, Khan M, Maslow C, Sandoval M, Mateu-Gelabert P, Krauss B, Aral SO (2008). Group sex events and HIV/STI risk in an urban network. *J. Acquir. Immune Defic. Syndr.* 49(4):440-6.

- Friedman SR, Bolyard M, Mateu-Gelabert P, Goltzman P, Pawlowicz MP, Singh DZ, Touze G, Rossi D, Maslow C, Sandoval M, Flom PL (2007). Some data-driven reflections on priorities in AIDS network research. *AIDS Behav.* 11(5):641-51.
- Friedman SR, Curtis R, Neaigus A, Jose B, Des Jarlais D (1999). *Social Networks, Drug Injectors' Lives, and HIV/AIDS*. NY: Springer.
- Friedman SR, Mateu-Gelabert P, Curtis R, Maslow C, Bolyard M, Sandoval M, Flom PL (2007). Social capital or networks, negotiations, and norms? A neighborhood case study. *Am. J. Prev. Med.* 32(6):160-170.
- Friedman SR, Neaigus A, Jose B, Curtis R, Goldstein M, Ildefonso G, Rothenberg RB, Des Jarlais DC (1997). Sociometric risk networks and risk for HIV infection. *Am. J. Public Health* 87(8):1289-96.
- Granovetter M (1973). The strength of weak ties. *Amer. J. Sociol.* 78(6):1360-80.
- Halfors DD, Iritani BJ, Miller WC, Bauer DJ (2007). Sexual and drug behavior patterns and HIV and STD racial disparities: the need for new directions. *Am. J. Public Health* 97(1):125-32.
- Heckathorn DD (1997). Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 44:174-99.
- Heckathorn DD, Semaan S, Broadhead RS, Hughes JJ (2002). Extensions of respondent-driven sampling: a new approach to the study of injection drug users aged 18-25. *AIDS Behav.* 6(1):55-67.
- Helleringer S, Kohler HP (2007). Sexual network structure and the spread of HIV in Africa: Evidence from Likoma Island, Malawi. *AIDS* 21(17):2323-32.
- Helleringer S, Kohler HP, Chimbiri A, Chatonda P, Mkandawire J (2009). The Likoma Network Study: Context, data collection, and initial results. *Demographic Res.* 21:427-68.
- Kelly JA, Amirkhani YA, Kabakchieva E, Vassileva S, McAuliffe TL, DiFranceisco WJ, Antonova R, Petrova E, Vassilev B, Khoursine RA, Dimitrov B (2006). Prevention of HIV and sexually transmitted diseases in high risk social networks of young Roma (Gypsy) men in Bulgaria: Randomised controlled trial. *BMJ* 333(7578):1098.
- Kendall C, Kerr LRF, Gondim RC, Werneck GL, Macena RHM, Pontes MK, Johnston LG, Sabin K, McFarland W (2008). An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. *AIDS Behav.* 12:97-104.
- Klovdahl AS, Potterat JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW (1994). Social networks and infectious disease: The Colorado Springs study. *Soc. Sci. Med.* 38(1):79-88.
- Kraut-Becher JR, Aral SO (2006). Patterns of age mixing and sexually transmitted infections. *Int. J. STD AIDS* 17(6):378-83.
- Latkin CA (1999). Outreach in natural settings: The use of peer leaders for HIV prevention among injecting drug users' networks. *Public Health* 113:151-9.
- Latkin CA, Forman V, Knowlton A, Sherman S (2003). Norms, social networks, and HIV-related risk behaviors among urban disadvantaged drug users. *Soc. Sci. Med.* 56(3):465-76.
- Lee SS, Tam DK, Tan Y, Mak WL, Wong KH, Chen JH, Yam WC (2009). An exploratory study on the social and genotypic clustering of HIV infection in men having sex with men. *AIDS* 23(13):1755-64.
- Lee SS, Tam DKP, Ho RLM, Wong KH (2009). Social network methodology for studying HIV epidemiology in men having sex with men. *J Infect Public Health* 2(4):177-83.
- Liebow E, McGrady G, Branch K, Vera M, Klovdahl A, Lovely R, Mueller C, Mann E (1995). Eliciting social network data and ecological model-building: focus on choice of name generators and administration of random-walk study procedures. *Soc Networks* 17(3-4):257-72.
- McGrady G, Marrow C, Myers G, Daniels M, Vera M, Liebow E, Klovdahl A, Lovely R (1995). A note on implementation of a random-walk design to study adolescent social networks. *Soc. Networks* 17(3-4):251-5.
- Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S (2009). Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. *Am. J. Public Health* 99(6):1023-31.
- Remple VP, Patrick DM, Johnston C, Tyndall MW, Jolly AM (2007). Clients of indoor commercial sex workers: heterogeneity in patronage patterns and implications for HIV and STI propagation through sexual networks. *Sex Transm. Dis.* 34(10):754-60.
- Rice E, Barman-Adhikari A, Milburn NG, Monro W (2012). Position-Specific HIV Risk in a Large Network of Homeless Youths. *Am. J. Public Health* 102(1):141-7.
- Rothenberg R, Baldwin J, Trotter R, Muth S (2001). The risk environment for HIV transmission: results from the Atlanta and Flagstaff network studies. *J. Urban Health* 78(3):419-32.
- Rothenberg RB, Long DM, Sterk CE, Pach A, Potterat JJ, Muth S, Baldwin JA, Trotter RT (2000). The Atlanta Urban Networks Study: a blueprint for endemic transmission. *AIDS* 14(14):2191-200.
- Rothenberg RB, Potterat JJ, Woodhouse DE (1996). Personal risk taking and the spread of disease: beyond core groups. *J Infect Dis.* 174 (2):144-9.
- Smith AMA, Grierson J, Wain D, Pitts M, Pattison P (2004). Associations between the sexual behaviour of men who have sex with men and the structure and composition of their social networks. *Sex Transm. Infect.* 80(6):455-8.
- Weeks MR, Clair S, Borgatti SP, Radda K, Schensul JJ (2002). Social networks of drug users in high-risk sites: Finding the connections. *AIDS Behav.* 6(2):193-206.
- Wejnert C (2009). Social network analysis with respondent-driven sampling data: A study of racial integration on campus. *Soc. Networks* 32(2):112-24.
- Williams CT, Latkin CA (2007). Neighborhood socioeconomic status, personal network attributes, and use of heroin and cocaine. *Am. J. Prev. Med.* 32(6):203-10.
- Woodhouse DE, Potterat JJ, Rothenberg RB, Darrow WW, Klovdahl AS, Muth SQ (1995). Ethical and legal issues in social network research: the real and the ideal. *NIDA Res. Monogr.* 151:131-43.
- Wylie JL, Shah L, Jolly A (2007). Incorporating geographic settings into a social network analysis of injection drug use and bloodborne pathogen prevalence. *Health Place* 13(3):617-28.