Downloaded from www.sacemaquarterly.com | The quarterly update on epidemiology from the South African Centre for Epidemiological Modelling and Analysis (SACEMA).

Published: September 2016

Connecting the dots: network data and models in HIV epidemiology

Wim Delva - Epidemiologist, SACEMA and Ghent University, Belgium.
Gabriel Leventhal - evolutionary biologist / theoretical ecologist at MIT and ETH Zurich.
Stéphane Helleringer- demographer / public health expert at Johns Hopkins University.

HIV is transmitted through social networks that are formed primarily by the sexual relationships and needle-sharing practices that people engage in. Therefore, while high-risk sexual behaviours of individuals, such as a large number of (concurrent) partners, and a high frequency of unprotected sex, increase their risk of acquiring HIV infection**,** individual-level factors alone are insufficient to fully explain the complex dynamics of HIV transmission. An individual's position in the sexual network also co-determines the probability of acquisition and onward transmission of HIV. For instance, a case-control study among pregnant women and their male partners in Lima, Peru, showed that the number of sexual relationships over the past five years of the male partners was predictive of the women's HIV infection status, independent of their own number of partners (1).

Network analysis in HIV epidemiology revolves around identifying HIV transmission pathways, i.e. the subsets of links across which HIV can spread. Some of these pathways are realised: they connect the people living with HIV (PLWH) that transmitted HIV to one another. Other pathways represent chains of potential transmission events: they link PLWH to individuals who are not (yet) infected, but are at risk of acquiring HIV in the future due to their network connections. Both empirical and modelling studies have provided evidence that the structure and dynamics of sexual networks shape the epidemiology of HIV infection (2, 3), and the success or failure of behavioural and biomedical interventions for HIV prevention (4, 5). Unfortunately, empirical data on sexual networks and on rates of HIV exposure and transmission within these networks are typically incomplete and unreliable because of feasibility challenges and social desirability bias (6).

Social scientists, molecular biologists and public health specialists have thus developed approaches to collecting partial or indirect network data, and subsequently infer HIV transmission pathways from such incomplete information. The social science approach uses data from behaviour and relationship surveys in combination with HIV testing to infer potential HIV transmission pathways (Figure 1C), whereas the molecular biology approach uses HIV genetic sequences from PLWH to identify realised HIV transmission pathways (Figure 1E). The public health approach seeks to identify high-risk networks by tracing and testing people connected to newly diagnosed HIV cases (Figure 1F).

The social science approach

The workhorse of the social science approach is the egocentric network survey. In this type of study, a random sample of individuals from the population of interest (Figure 1C) is asked to provide information about their recent sexual and/or needle sharing partners (e.g. their age, race and sex), and to describe the characteristics of these relationships (e.g. start and end dates, condom use). Respondents may also be invited to test for HIV infection. Egocentric surveys permit measuring characteristics of the personal networks of respondents such as homophily (the propensity to engage in partnerships with others who share similar characteristics), or concurrency (the likelihood of having more than one ongoing relationship at one point in time). But they do not provide data on HIV transmission chains because the partners of sampled respondents are not typically enrolled in the study. For this reason, potential HIV transmission pathways can only be inferred. Network inferences have greatly improved recently thanks to the development of exponential random graph models (ERGMs). ERGMs are a family of statistical models that can accommodate the interdependencies between individuals that characterise network datasets (8). They were originally developed for the analysis of complete network datasets in which all individuals and links are listed. But they can also be used with incomplete data from egocentric studies under certain simplifying assumptions.

Naturally, the accuracy of inferences about HIV transmission pathways derived from egocentric data depends on the validity of the model of network formation. For example, important groups of individuals such as mobile and marginalised key populations (sex workers, injecting drug users) may be underrepresented in egocentric surveys, and PLWH who are aware of their status may be significantly less likely to participate in surveys that include HIV testing. Another limitation stems from egocentric data providing intrinsically incomplete and often inaccurate data on the links that connect individuals in a population. In order to minimise respondent fatigue and recall errors, egocentric surveys often only elicit responses for a small number of relationships per respondent, e.g. their three or five most recent relations. The reported personal networks are thus likely truncated. Furthermore, because of recall bias or social desirability bias, survey respondents often omit to report some of their relationships during interviews and misreport the characteristics of some of their partners. These various forms of missing data affect network inferences from egocentric survey data.

The molecular biology approach

Whereas social scientists often start from a random sample of individuals irrespective of their HIV serostatus, the molecular biology approach focuses on PLWH. For HIV, as for many other retroviruses, the rate at which viral populations undergo genetic changes within each HIV-positive person (or "host") is orders of magnitude faster than the rate at which they are transmitted between hosts (9). These genetic differences between viral populations in different hosts can be used to infer the most likely evolutionary history of the pathogen.

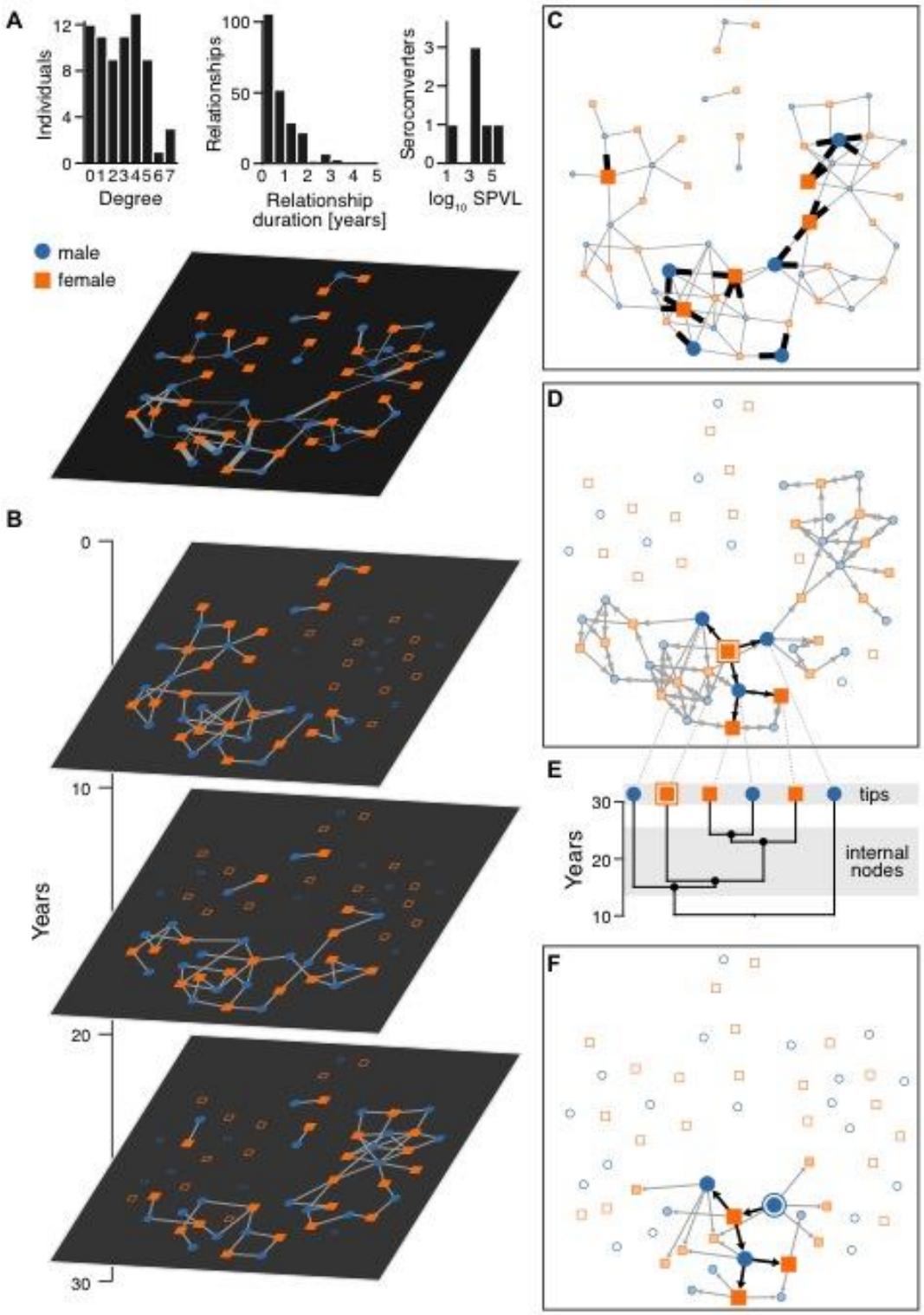

Figure 1. Graphical representations of simulated networks of sexual relationships and HIV transmission.

Notes: All data from Figure 1 are synthetic and were generated using Simpact, a freely available agent-based modelling tool for simulating HIV transmission in dynamic sexual networks (7). ***A.*** *Complete (cumulative) network of all sexual relationships that were formed over a 30-year time period between members of the simulated population. Thicker edges represent longer relationships. Histograms at the top of the panel A represent the degree distribution (lifetime number of partners) in the population, the distribution of relationship durations, and the distribution of log10 set-point viral load (SPVL) among people living with HIV.* ***B.*** *Each graph represents sexual relationships that existed in the first, middle and last 10-year time slice.* ***C.*** *Simulated egocentric network survey conducted in the population represented in panel A. Larger nodes are participants in the survey and the black edges represent their reported relationships in the last 10 years prior to the survey at the end of the 30-year simulation period.* ***D.*** *In grey, the potential transmission pathways resulting from the introduction of HIV into the simulated population from the index case (framed square). The black edges represent the realised HIV transmission pathway, connecting individuals that were infected with HIV by the end of the simulation period.* ***E.*** *Phylogenetic tree, reconstructed from virus samples.* ***F.*** *Simulated contact tracing investigation starting from the framed circle, i.e. the first person to present to a clinic and be diagnosed with HIV in this simulated population. Arrows indicate relationships reported during contact tracing interviews. Black arrows connect individuals that were seropositive after testing. Grey edges, on the other hand, represent reported relationships in which the partner was not infected with HIV.*

The molecular study of networks then entails obtaining viral sequences from PLWH (10), and grouping HIV sequences by genetic similarity (11). These groupings are depicted as phylogenetic trees (or phylogenies), where tips in the tree represent PLWH, the branching pattern indicates the genetic similarity between sequences from different PLWH, and the internal nodes of the tree represent past transmission events (Figure 1E).

Phylogenies have been used to identify HIV transmission clusters, i.e. groups of PLWH with highly similar viral populations and who are likely connected by an HIV transmission pathway. When cluster analysis incorporates the demographic, behavioural and clinical characteristics of cluster members, this information may help guide the targeting of HIV prevention and treatment programmes. Viral linkage analysis in HIV clinical trials with serodiscordant couples has enabled more accurate estimation of the efficacy of early ART (12) and genital herpes suppression (13) to prevent HIV transmission.

The shape of a phylogenetic tree (also called its "topology") can help elucidate aspects of the broader structure of the social network and HIV transmission pathways within them (Figure 2). However, different transmission networks can yield phylogenetic trees with similar topologies. Additional summary tree statistics, such as branch lengths, tree width, tree depth, and the occurrence of "cherries" and "ladders" (Figure 2C), are then needed to differentiate between homogeneous, chain-like and super-spreader transmission networks (Figure 2B) (14).

An important advantage of phylogenetic analyses over interview-based methods is that they are not subject to recall or social desirability bias. And unlike egocentric studies, they elicit indirect connections between individuals who may be at two or more degrees of separation. Molecular biology thus allows longer-range investigations of the connectivity of networks, including across geographical regions, age and racial/ethnic groups, and between subpopulations.

A major limitation of the molecular biology approach to network inference is that it requires a high sampling density (i.e. the proportion of PLWH for whom a viral sequence is available. If it is too low, then it will be difficult to link PLWH to the source of their infection, and the phylogenetic tree will be sparser than it really is.

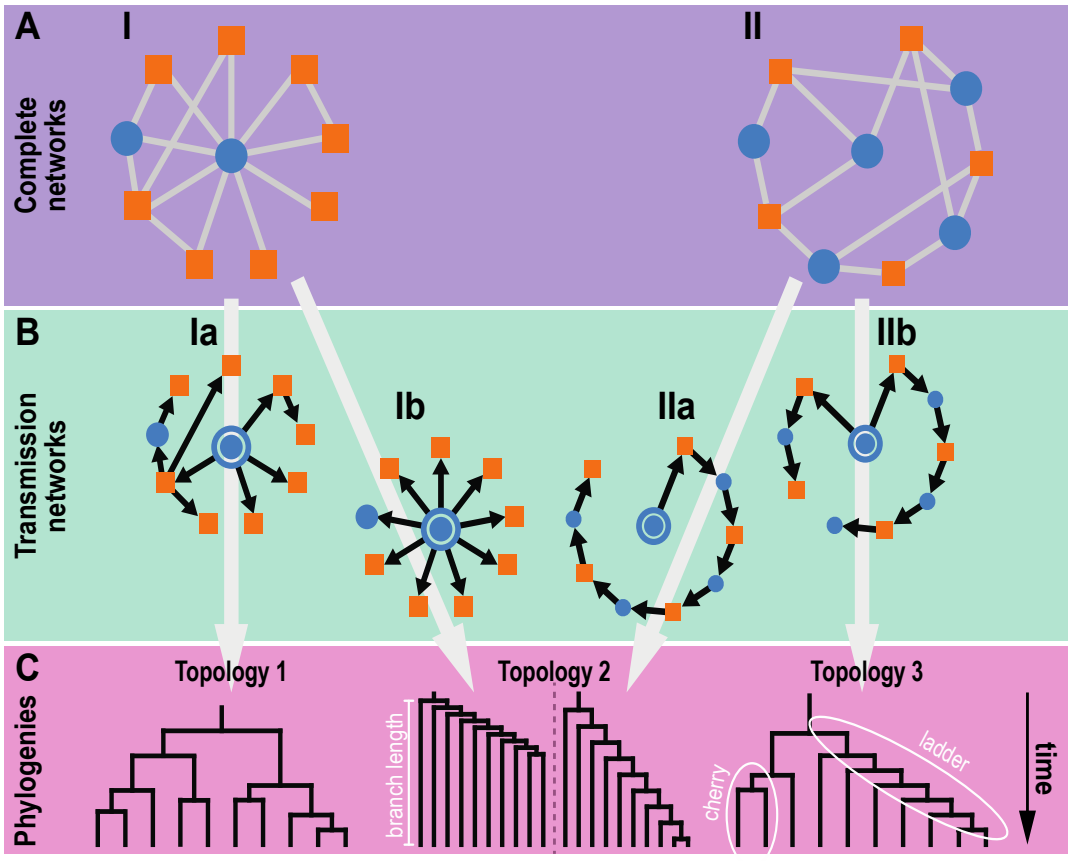

Fig. 2. The complete network influences the realized transmission pathway and subsequent phylogenetic tree, but only to a certain degree.

Notes: ***A.*** *Two examples of differing complete networks.* ***B.*** *Many different HIV transmission pathways may emerge following the infection of a single individual (framed circle). Individual characteristics such as sex (circles/squares) can restrict the possible transmission networks, but mixed modes of transmission (e.g. heterosexual, MSM, and needle sharing) further extend the range of possible transmission networks.* ***C.*** *The transmission networks from (b) leave a trace in the phylogenetic tree topology. The topology of a phylogenetic tree is defined as the structure of the tree with its tips (or 'leaves', i.e. the end points of the tree), without paying attention to branch lengths and left-right ordering of branches and leaves. Topologies can be compared using imbalance (topologies 2 and 3 are less balanced, i.e. more asymmetrical than topology 1), but different transmission networks (Ib and IIa) may result in the same topology (topology 2). In these cases, additional information such as branch lengths and the number of cherries and ladders may help differentiate between transmission networks.*

A recent study in Botswana suggested that a sampling density of 50-70% was required for accurate identification of transmission clusters (15). This is particularly problematic for populations with high HIV prevalence and incidence, where a prohibitively large number of HIV sequences may be needed to achieve an appropriately high sampling density. Furthermore, phylogenetic methods do not distinguish between transmission events that occurred through sexual intercourse, or through sharing of infected injection equipment, thus limiting our ability to disentangle the contributions of different modes of transmission in bridging populations that concurrently engage in multiple high-risk practices.

The public health approach

Network data can also be generated when HIV treatment and prevention programmes offer HIV partner notification (PN) services to PLWH. In PN, newly diagnosed PLWH are asked to provide a list of their recent sexual or injection partners, along with contact details, so that these partners can be traced. Identifiable partners are then contacted and visited in person by a disease notification specialist. They are informed about their potential exposure to HIV, and are invited to visit a health facility for HIV testing and linkage to care if indicated. The process of contact solicitation, contact tracing and testing is repeated for notified partners who test positive for HIV, but it stops for partners who are HIV-negative.

PN may help uncover subsets of HIV transmission pathways that often remain hidden in egocentric or phylogenetic studies. These include hard-to-reach high-risk groups (16), stigmatised populations that rarely attend health facilities and mobile individuals who are less likely to be included in traditional survey sampling frames. PN is more effective at reaching such groups because its network sampling process is "adaptive" in the sense that it uses information provided directly by network members to guide the selection and recruitment of individuals.

The downside of this adaptive sampling process is that it can also be highly selective. Not every newly diagnosed PLWH will choose to notify their partners and the decision to use PN services may be related to recent risk behaviours. PLWH may deliberately choose not to mention some of their recent partners during PN interviews, or may not recall sufficient details about other partners to enable PN. And even among those partners who are sought out by the disease notification specialist, some may never be successfully contacted due to insufficient information, while others may reject PN. It is therefore unclear which parts of the HIV transmission chains the PN process reveals (17). In regions with high HIV prevalence, offering PN requires significant investments in health personnel, record keeping and data linkages, which can be prohibitive in low-resource settings.

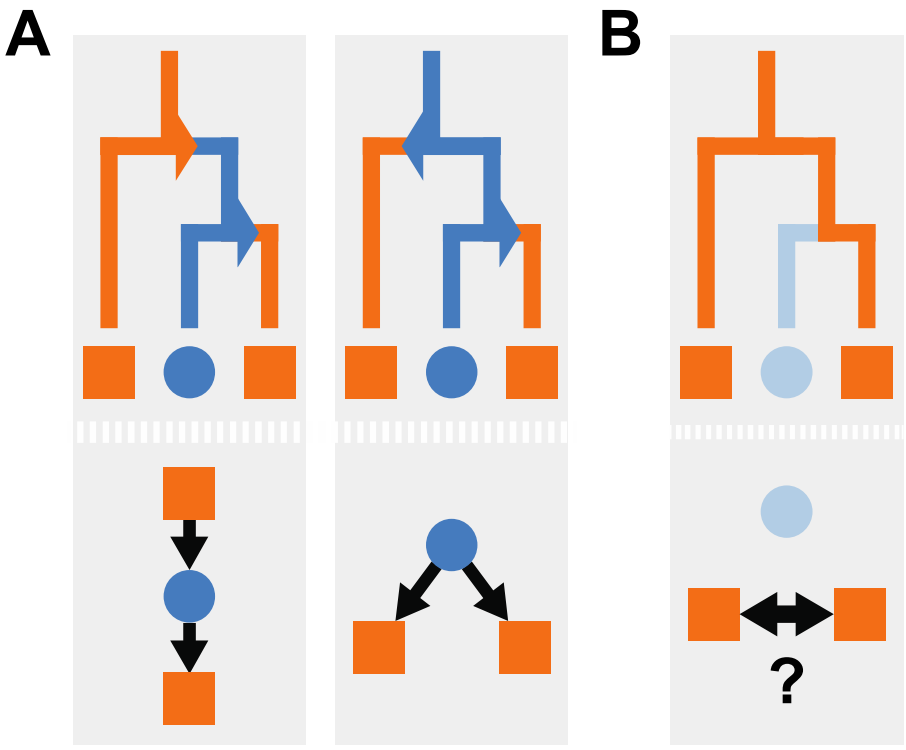

Figure 3: Phylogenies provide incomplete transmission network data.

Notes: ***A.*** *The topology of a phylogenetic tree is defined as the structure of the tree with its tips ("leaves", i.e. the end points of the tree), without paying attention to branch lengths and left-right ordering of branches and leaves. The topology of the tree alone, however, cannot resolve who infected whom. For example, the topology of the tree represented in the upper panels of Figure 3A is compatible with multiple directed networks (lower panels of Figure 3A).* ***B.*** *Incomplete sampling of PLWH may also result in incorrect identification of transmission pairs in phylogenetic trees. In this example, if the node represented by the circle is not sampled, we may incorrectly infer transmission between the two squares, whereas the transmission actually occurred through a longer chain.*

In conclusion, data on sexual networks come from an increasingly diverse array of sources, but each of these sources only document parts of the networks through which HIV may spread. Egocentric network surveys suffer from non-response, social desirability bias and the inability to probe beyond the immediate network connections of individuals. Through partner notification services, realised and potential HIV transmission pathways may be partially revealed, but in resource-poor settings with generalised HIV epidemics offering this may require prohibitively large investments. Phylogenetic tree analysis permits reconstructing parts of the HIV transmission chains by linking genetically related infections, but to be informative, HIV sequence data must be available for what may be an unfeasibly large sample of PLWH. Novel methods to combine these data sources are beginning to emerge from the collaborative efforts of experts in computational biology, social science, statistics, public health and epidemiological modelling. Further advances in network analysis for HIV epidemiology will require (1) important methodological developments in network modelling, as well as (2) a long-term, global commitment from researchers and funding agencies to ensure open access to analytical tools and multifaceted network datasets that include HIV sequences along with behavioural, demographic, clinical and programmatic information.

Wim Delva *- Epidemiologist, SACEMA and Ghent University, Belgium. Areas of interest: statistical analysis of sexual behaviour data, stochastic and deterministic modelling of sexual network dynamics and HIV transmission. Wim.Delva@ugent.be*

Gabriel Leventhal *- evolutionary biologist / theoretical ecologist at MIT and ETH Zurich. Areas of interest: phylodynamics of infectious diseases, ecological networks, heritability of HIV set point viral load. gaberoo@mit.edu*

Stéphane Helleringer *- demographer / public health expert at Johns Hopkins University. Areas of interest: HIV, Ebola, network epidemiology, measurement of mortality and causes of death. sheller7@jhu.edu*

Acknowledgements: We thank Jori Liesenborgs, who developed and documented the Simpact Cyan program (C++), as well as the RSimpactCyan front end (R package), and Gavin Hitchcock, Roxanne Beauclair and Niel Hens for their helpful comments on an earlier version of this paper.

References:

1. Johnson KM, Alarcon J, Watts DM, Rodriguez C, Velasquez C, Sanchez J, et al. Sexual networks of pregnant women with and without HIV infection. AIDS. 2003;17:605-612.
2. Morris M, Epstein H, Wawer M. Timing is everything: international variations in historical sexual partnership concurrency and HIV prevalence. PLoS One. 2010;5:e14092.
3. Helleringer S, Kohler HP. Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi. AIDS. 2007;21:2323-2332.
4. Hontelez JA, Lurie MN, Barnighausen T, Bakker R, Baltussen R, Tanser F, et al. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study. PLoS Med. 2013;10:e1001534.
5. Hurt CB, Beagle S, Leone PA, Sugarbaker A, Pike E, Kuruc J, et al. Investigating a sexual network of black men who have sex with men: implications for transmission and prevention of HIV infection in the United States. J Acquir Immune Defic Syndr. 2012;61:515-521.
6. Crawford M, Popp D. Sexual double standards: a review and methodological critique of two decades of research. J Sex Res. 2003;40:13-26.
7. Liesenborgs J, Meng F, Hens N, Delva W. Simpact Cyan 0.19.4. [http://www.simpact.org/how-simpact-works/] 2015. Accessed 14 June 2016.
8. Chatterjee S, Diaconis P. Estimating and understanding exponential random graph models. The Annals of Statistics. 2013;41:2428-2461.
9. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet. 2004;5:52-61.
10. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS. 2014;9:126-133.
11. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. AIDS Rev. 2006;8:125-140.
12. Eshleman SH, Hudelson SE, Redd AD, Wang L, Debes R, Chen YQ, et al. Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial. J Infect Dis. 2011;204:1918-1926.
13. Campbell MS, Mullins JI, Hughes JP, Celum C, Wong KG, Raugi DN, et al. Viral linkage in HIV-1 seroconverters and their partners in an HIV-1 prevention clinical trial. PLoS One. 2011;6:e16986.
14. Colijn C, Gardy J. Phylogenetic tree shapes resolve disease transmission patterns. Evol Med Public Health. 2014:2014(1):96-108.
15. Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. Impact of sampling density on the extent of HIV clustering. AIDS Res Hum Retroviruses. 2014;30:1226-1235.
16. Adams J, Moody J, Muth SQ, Morris M. Quantifying the Benefits of Link-Tracing Designs for Partnership Network Studies. Field Methods. 2012;24:175-193.
17. Begley EB, Oster AM, Song B, Lesondak L, Voorhees K, Esquivel M, et al. Incorporating rapid HIV testing into partner counseling and referral services. Public Health Rep. 2008;123 Suppl 3:126-135.