Published in final edited form as:

Ann Epidemiol. 2014 July ; 24(7): 554–557. doi:10.1016/j.annepidem.2014.04.006.

Using Sociometric Measures to Assess Non-Response Bias

Britt Livak, MPH¹ and John A. Schneider, MD, MPH^{1,2}

¹Department of Health Studies, 5841 South Maryland Avenue, University of Chicago, Chicago IL 60637

²Department of Medicine, 5841 South Maryland Avenue, University of Chicago, Chicago IL 60637

Abstract

Purpose—Much attention has been given to the potential non-response bias that occurs in epidemiologic studies that attempt to enroll a representative sample. Most analyses surrounding non-respondents focus on individual-level attributes and how they vary across respondents and non-respondents. While these attributes are of interest, analysis of the social network position of non-respondents as defined by traditional sociometric measures (i.e. centrality, bridging) has not been conducted, and could provide further insights into the validity of the sample.

Methods—We utilized data from the Secunderabadi Mens' Study, a whole network of Indian men who have sex with men (MSM) generated using cell phone contact lists of men approached using Time Location Cluster Sampling. Multivariable logistic regression was used to determine whether demographic and behavioral attributes and in-degree (the frequency that a MSM was listed across all cell phone contact lists) were associated with being a respondent.

Results—239 respondents were interviewed and 81 were approached but did not consent to the interview ("non-respondents").

Conclusions—Respondents were more likely to have higher in-degree than non-respondents, adjusting for attribute differences (OR 1.19; 95% CI 1.07, 1.34). This analysis suggests that the network position of non-respondents may be important when considering the potential impact of non-response bias.

MeSH Headings

Epidemiologic Biases; Social Networking; Data Collection

INTRODUCTION

Participation in epidemiologic studies has been declining in recent years, and this non-participation may significantly bias the interpretation of study results.(1) Much attention has

© 2014 Elsevier Inc. All rights reserved.

Corresponding author: Britt Livak, MPH, Departments of Health Studies, 5841 South Maryland Avenue MC 2007, University of Chicago, IL 60637, 773-835-4004, brittsl@uchicago.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

been given to the characterization of individuals who are and are not likely to participate in epidemiologic studies as a way to assess the potential bias. However, most analyses surrounding non-participation bias have focused upon individual-level attributes, such as demographics, health status, and exposure to risk factors.(1) While individual-level attributes are of interest in assessing the representativeness of the sample generated, analyses of attributes related to the position of individuals in their social networks, referred to as sociometric measures, also have implications for determining sample representativeness. Sociometric measures can include the number of connections one person has to other people “degree centrality”, the extent to which an individual is close to other individuals in a network “closeness”, and the extent to which a person is connected to other people who are not connected to each other “betweenness”, for example.(2) These measures are necessary to consider because they impact an individual’s behavior, which has consequences for their health.(3)

Social network analysis provides critical insights into the dynamics of health outcomes. Social network data can be used to intervene on disease transmission, for instance through contact tracing methods such as those used to cease the cholera epidemic. Proliferation of HIV negative components of small network size, for instance, has been found to contribute to the stabilization of HIV prevalence among injection drug users in New York City.(4) It has also been established that social norms within a network affect health outcomes by influencing health behaviors.(5, 6) Peer norms have been associated with HIV risk behaviors among drug users as well as men who have sex with men (MSM).(5, 7) Network influences pertain to diseases other than HIV as well. The type of contacts that an individual has in their network, such as the proportion of intimate ties in a network, has also been linked to reduced risk of cardiovascular disease, and tobacco use among adolescents.(8) It is therefore important to insure that these sociometric measures are captured and representative in the study sample, as are other traditionally collected individual level attributes such as age, gender and education.

Respondent Driven Sampling (RDS) is a recruitment methodology that is commonly used in network studies, particularly for hard-to-reach populations.(9) RDS is a modified form of chain referral where investigators select initial respondents, “seeds” to recruit their confidants into the study. Recruitment bias associated with the use of this methodology has been assessed previously. Post-hoc analytic methods have been developed to control for recruits’ network size and likelihood to recruit other participants who share the same socio-demographic characteristics. (10) However, similar assessments have not been made using other types of network recruitment methodology.

We use data from a network generated with Cell-phone Assisted Network Detection and Identification technology (11) to objectively compare sociometric measures between respondents and non-respondents. Using objective network data allows us to avoid issues related to individuals over or underreporting their social interactions, otherwise known as “expansiveness bias.”(12) Sociometric measures, such as centrality (2), are rarely assessed in traditional epidemiology studies to determine sample representativeness. Our unique approach provides an opportunity to objectively compare network-level characteristics of

respondents to non-respondents in order to further advance the assessment for nonresponse bias in epidemiologic studies.

METHODS

Data for this analysis come from the Secunderabadi Mens' Survey. The study took place in a large city in Southern India at 20 social venues where men who have sex with men (MSM) congregate to socialize. (11) The study population consisted of individuals identifying as male who: were at least 18 years of age, visit one of the 20 venues, reported anal/oral intercourse with another man within the previous 12 months, and owned and were in possession of at least one cell-phone at the time of recruitment. Individuals who shared cell-phones were ineligible.

Data were collected using Time Location Cluster Sampling. Every month, 15 venues were randomly selected (without replacement) from the sampling frame that included all venues. Three-hour periods associated with the venue were then randomly selected, and MSM were randomly approached at each venue. MSM were approached at the venues and evaluated for eligibility by the criteria described above. Limited socio-demographic questions were asked of men who refused to participate. All study participants were surveyed about relationship characteristics and demographic characteristics, as well as provided dry blood spots for HIV testing.

The MSM network was assembled by using a Subscriber Identity Module card reader and associated software to extract consecutive respondents' contact lists. Contact lists of all sampled respondents were linked using the cell phone number as a unique identifier to generate an "*augmented*" communication network. Information on contact list network members was collected, including whether their contacts were MSM. The network analysis was restricted to MSM contacts. Respondents were asked about each contact in their cell phone with regard to the type of relationship they have with each person (e.g. friend, sex partner, relative, etc.). Provision of phone numbers allowed for assessment of non-respondents sociometric positioning within the MSM network.

Sampling occurred between July 1st and December 31st of 2010. We recruited individuals until we reached network saturation in the region (until there was a 95% likelihood that each subsequent recruit would already have been linked in the network through another participant's contact list). We used a network redundancy curve fit from data on index of respondents and week of respondent interviews versus network size to exponential model to determine network saturation. The data was fit to a scaled/shifted exponential cumulative distribution function $f(x)=99.2-95.9e^{(-4.9x)}$ where x represents the index of the respondent and $f(x)$ represents network size. This permitted us to compare sociometric measures between respondents and non-respondents because it allowed for both respondents and non-respondents to be equally likely to be included in the network. All procedures were approved by Institutional Ethics Committees at the University of Chicago in the United States and SHARE-India in India.

Measures

The following individual attributes were assessed for each MSM: Marital Status, Caste (Backward Caste, Scheduled Caste, Scheduled Tribe, Other Caste), Religion (e.g. Hindu, Muslim, Christian, Sikh), MSM sex position identity (e.g. receptive, insertive, versatile) and participation in exchange sex in the three months prior to the survey. All eligible individuals were contacted up to three times for recruitment into the study. Respondents were defined as MSM who completed the survey, and non-respondents were defined as individuals who were approached and eligible, but did not consent to the survey. In-degree centrality, one of the most commonly cited network metrics, was used for each MSM. In this instance it is a measure of the number of times an MSM was listed across all cell phone contact lists in the sample and thus signifies how centrally located they are within the network.⁽²⁾ In-degree was chosen rather than other sociometric measures such as out-degree to avoid bias due to the lack of information about the networks of non-respondents. In other words, the augmented communication network allowed us to position non-respondents within the network, but we could not assume the direction of communication between non-respondents and their alters.

Analysis

Chi-square tests and Mann-Whitney *u* tests were used to detect differences between demographic attributes, in-degree, and respondent status. We examined whether demographic attributes and in-degree were associated with being a respondent using bivariable and multivariable logistic regression. Variables with two-sided $P < 0.20$ in bivariable analysis were considered as candidates for multivariable models, with two-sided $P < 0.05$ as the cutoff for retention in the final models.

RESULTS

A total of 239 respondents were interviewed and an additional 81 individuals were approached, but did not consent to the interview. The main reasons for non-participation were that the individual was not reachable following initial contact (41%), that the client was too busy to participate (39%), and that the client was out of town (7%).

The majority of respondents were not married (91%), of a lower caste (82%), Hindu (91%), and reported having exchange sex (56%). Respondents identified with being primarily receptive MSM position (39%) more commonly than they identified as being insertive (33%) or versatile (28%). The mean number of in-degree connections among respondents was 3.5 (standard deviation (SD), 2.8). The majority of non-respondents were not married (72%), of lower caste (78%), reported having exchange sex (64%), and approximately half were Sikh (47%). The majority of non-respondents identified with being primarily receptive MSM position (54%), with 20% being insertive, and 26% being versatile. The mean number of in-degree connections among non-respondents was 2.8 (SD 3.3). Respondents differed from non-respondents in bivariate analyses on marital status ($P < 0.0001$), caste ($P = 0.001$), religion ($P < 0.0001$), and MSM position ($P = 0.03$). They did not differ on in-degree ($P = 0.23$) or exchange sex ($P = 0.22$).

Figure 1 displays the network of respondents and non-respondents by in-degree. In-degree is depicted by node size, with larger nodes indicating larger in-degree. The unadjusted mean in-degree was similar among respondents and non-respondents, as can be seen by the similar distribution of different sized nodes in each group in figure 1. The difference was not statistically significant in the bivariate analysis ($P=.23$).

Table 1 displays the results of the multivariable analysis. After adjusting for differences in individual attributes in multivariable logistic regression, the association between in-degree and respondent status strengthened. For every one person increase in network size (in-degree), the odds of being a respondent increased by 19% (OR 1.19; 95% CI 1.07, 1.34).

In addition to in-degree, the odds of being a respondent among men who identify as the receptive MSM position were 66% lower than the odds of being a respondent among men who identify as the insertive MSM position (OR 0.44; 95% CI 0.19, 0.99). The odds of being a respondent among the married men were 80% less than the odds of being a respondent among unmarried men (OR 0.19; 95% CI 0.19, 0.99), and the odds of being a respondent among men with non-Hindu religious identity were 94% less than the odds of being a respondent among men with Hindu religious identity (OR 0.06; 95% CI 0.03, 0.12).

CONCLUSION

We found a significant difference in in-degree between respondents and non-respondents after controlling for differences in individual attributes. This analysis suggests that the network position of non-respondents may be important when considering the potential impact of non-response bias. Ignoring sociometric measures may lead to biased conclusions because it assumes that individuals act independently from individuals in their networks. Previous literature has also shown that participant non-response may bias the calculation of network-based statistics. Specifically, non-response may lead to underestimates of the calculation of the tendencies of individuals to connect with others that are similar to themselves (assortativity). (13)

This analysis was limited in that sociometric measures other than in-degree could not be assessed due to biases associated with the lack of information about the networks of non-respondents. Some of the individual attribute categories had to be collapsed in the multivariable analysis due to small cell size, which limited the amount of detail provided about these variables. Lastly, the data were not collected as a random sample. Therefore, the results cannot be interpreted as being representative of all MSM in southern India.

Our findings have implications for a number of fields. Aside from HIV, social network analysis has been used to assess the network drivers of substance abuse (14), suicide (15), smoking (16), obesity (17), romantic relationships (18), contraceptive use (19, 20), and physician behavior.(21),(3) Non-response bias should be assessed in terms of sociometric metrics in these fields as well. Our results indicate that after controlling for differences in individual attributes, those who are objectively more socially isolated may have a decreased probability of being selected for research studies. Recruitment strategies that target socially

isolated individuals, perhaps via web-based modes of data collection, should be developed to mitigate this bias.

Acknowledgments

This work received funding from the National Institutes of Health grants R21HD068352 and R21AI098599.

Abbreviations

MSM	Men who have sex with men
CI	Confidence Interval
OR	Odds Ratio

References

1. Galea S, Tracy M. Participation rates in epidemiologic studies. 2007;1047–2797. (Print). eng.
2. Freeman LC. Centrality in social networks. Conceptual clarification. *Social networks*. 1979; (1): 215–39.
3. Valente, TW. *Social Networks and Health: Models, Methods, and Applications*. New York, New York: Oxford University Press; 2010.
4. Friedman SR, Kottiri BJ, Neaigus A, Curtis R, Vermund SH, Des Jarlais DC. Network-related Mechanisms May Help Explain Long-term HIV-1 Seroprevalence Levels That Remain High but Do Not Approach Population-Group Saturation. *American Journal of Epidemiology*. 2000 Nov 15; 152(10):913–22. [PubMed: 11092433]
5. Latkin, CA.; Forman, V.; Knowlton, A.; Sherman, S. Norms, social networks, and HIV-related risk behaviors among urban disadvantaged drug users. 2003.
6. Valente TW. Network Interventions. *Science*. 2012; 337(49)
7. Schneider, JCB.; Ostrow, DMS.; Schumm, P.; Laumann, EOFS. Network Mixing and Network Influences Most Linked to HIV Infection and Risk Behavior in the HIV Epidemic Among Black Men Who Have Sex With Men. 2012.
8. Barefoot JC, Grønbaek M, Jensen G, Schnohr P, Prescott E. Social Network Diversity and Risks of Ischemic Heart Disease and Total Mortality: Findings from the Copenhagen City Heart Study. *American Journal of Epidemiology*. 2005 May 15; 161(10):960–7. [PubMed: 15870160]
9. Heckathorn, DD. *Social Problems*. 1997. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations.
10. Heckathorn, DD. *Sociological Methodology*. 2007. Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment.
11. Schneider, JAZA.; Laumann, EO. *Social Science and Medicine*. A New HIV Prevention Network Approach: Sociometric Peer Change Agent Selection. In Press
12. Feld SL, Carter WC. Detecting measurement bias in respondent reports of personal networks. *Social networks*. 2002; 24(4):365–83.
13. Kossinets G. Effects of missing data in social networks. *Social networks*. 2006; 28:247–68.
14. Valente TW, Mouttapa M, Gallaher M. Social network analysis for understanding substance abuse: A transdisciplinary perspective. *Substance Use & Misuse*. 2004; 39:1685–712. [PubMed: 15587948]
15. Bearman PS, Moody J. Suicide and friendships among American adolescents. *American Journal of Public Health*. 2004; 94:89–94. [PubMed: 14713704]
16. Alexander C, Piazza M, Mekos D, Valente TW. Peer networks and adolescent cigarette smoking: An analysis of the national longitudinal study of adolescent health. *Journal of Adolescent Health*. 2001; (29):22–30. [PubMed: 11429302]

17. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*. 2007; 357:370–9. [PubMed: 17652652]
18. Bearman PS, Moody J, Stovel K. Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*. 110:44–91.
19. Entwisle B, Rindfuss RD, Guilkey DK, Chamrathirong A, Curran SR, Sawangdee Y. Community and contraceptive choice in rural Thailand: A case study of Nang Rong. *Demography*. 1996; 33:1–11. [PubMed: 8690134]
20. Valente TW, Watkins S, Jato MN, Van der Straten A, Tsitsol LM. Social network associations with contraceptive use among Cameroonian women in voluntary associations. *Social Science and Medicine*. 1997; 45:677–87. [PubMed: 9226791]
21. Gross CP, Cruz-Correa M, Canto MI, McNeil-Solis C, Valente TW, Powe NR. The adoption of ablation therapy for Barrett's esophagus: A cohort study of gastroenterologists. *American Journal of Gastroenterology*. 2002; 97:279–86. [PubMed: 11866262]

Figure 1.

Network of Respondents and Non-Respondents by In-degree

*Node size reflects in-degree, with larger nodes have larger in-degree values.

*Graph does not display contacts that were named but not approached for survey participation.

Table 1

Multivariable Logistic Regression of In-Degree and Respondent Status adjusting for Individual Respondent Characteristics; Secunderabadi Mens' Survey ^a

	Odds Ratio	95% CI	P-value
In-Degree	1.19	1.07, 1.34	0.002
MSM Type			
Insertive	Ref.		
Receptive	0.44	0.19, 0.99	0.05
Versatile	0.61	0.24, 1.57	0.31
Married			
No	Ref.		
Yes	0.19	0.19, 0.99	<0.0001
Religion			
Hindu	Ref.		
Other ^b	0.06	0.03, 0.12	<0.0001

OR, Odds Ratio; CI, Confidence Interval;

^a Model adjusts for Caste, Exchange Sex, and all variables present in the table,

^b Includes Muslim, Christian, Sikh