Improving the Network Scale-Up Estimator: Incorporating Means of Sums, Recursive Back Estimation, and Sampling Weights

Patrick Habecker*, Kirk Dombrowski, Bilal Khan

Department of Sociology, University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America

* phabecker2@unl.edu

Abstract

Researchers interested in studying populations that are difficult to reach through traditional survey methods can now draw on a range of methods to access these populations. Yet many of these methods are more expensive and difficult to implement than studies using conventional sampling frames and trusted sampling methods. The network scale-up method (NSUM) provides a middle ground for researchers who wish to estimate the size of a hidden population, but lack the resources to conduct a more specialized hidden population study. Through this method it is possible to generate population estimates for a wide variety of groups that are perhaps unwilling to self-identify as such (for example, users of illegal drugs or other stigmatized populations) via traditional survey tools such as telephone or mail surveys—by asking a representative sample to estimate the number of people they know who are members of such a "hidden" subpopulation. The original estimator is formulated to minimize the weight a single scaling variable can exert upon the estimates. We argue that this introduces hidden and difficult to predict biases, and instead propose a series of methodological advances on the traditional scale-up estimation procedure, including a new estimator. Additionally, we formalize the incorporation of sample weights into the network scale-up estimation process, and propose a recursive process of back estimation "trimming" to identify and remove poorly performing predictors from the estimation process. To demonstrate these suggestions we use data from a network scale-up mail survey conducted in Nebraska during 2014. We find that using the new estimator and recursive trimming process provides more accurate estimates, especially when used in conjunction with sampling weights.

Introduction

Due to the difficulty of studying hidden and hard-to reach populations, many researchers have moved past large general surveys to develop a specialized set of data collection methods. These techniques include observation, participation, key informant interviews, and location-based strategies that can provide valuable insight about population characteristics, but which rarely generate measures of representativeness of the sample or the size of the population as a whole.

6. Gile KJ, Handcock MS. Respondent-driven sampling: an assessment of current methodology. Sociol Methodol. 2010; 40(1):285–327. PMID: 22969167

7. Handcock MS, Gile KJ. Modeling social networks from sampled data. Ann Appl Stat. 2010; 4(1):5–25. PMID: 26561513

8. Killworth PD, McCarty C, Bernard HR, Shelley GA, Johnsen EC. Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. Eval Rev. 1998; 22:289–308. PMID: 10183307

9. Shokoohi M, Baneshi MR, Haghdoost A-A. Size estimation of groups at high risk of HIV/AIDS using network scale up in Kerman, Iran. Int J Prev Med. 2012 Jul; 3(7):471–6. PMID: 22891148

10. Russell Bernard H, Johnsen EC, Killworth PD, Robinson S. Estimating the size of an average personal network and of an event subpopulation: some empirical results. Soc Sci Res. 1991 Jun; 20(2):109–21.

11. Ezoe S, Morooka T, Noda T, Sabin ML, Koike S. Population size estimation of men who have sex with men through the network scale-up method in Japan. PLoS ONE. 2012 Jan 27; 7(1):e31184. doi: 10. 1371/journal.pone.0031184 PMID: 22563366

12. Guo W, Bao S, Lin W, Wu G, Zhang W, Hladik W, et al. Estimating the size of HIV key affected populations in Chongqing, China, using the network scale-up method. PLoS ONE. 2013 Aug 13; 8(8):e71796. doi: 10.1371/journal.pone.0071796 PMID: 23967246

13. Bernard HR, Hallett T, Iovita A, Johnsen EC, Lyerla R, McCarty C, et al. Counting hard-to-count populations: the network scale-up method for public health. Sex Transm Infect. 2010 Dec 1; 86(Suppl 2):ii11–5. doi: 10.1136/sti.2010.044446 PMID: 21106509

14. McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA. Comparing two methods for estimating network size. Hum Organ. 2001 Mar 1; 60(1):28–39.

15. Johnsen EC, Bernard HR, Killworth PD, Shelley GA, McCarty C. A social network approach to corroborating the number of AIDS/HIV + victims in the US. Soc Netw. 1995 Jul; 17(3–4):167–87.

16. Iannacchione VG. The changing role of address-based sampling in survey research. Public Opin Q. 2011 Sep 1; 75(3):556–75.

17. Link MW, Battaglia MP, Frankel MR, Osborn L, Mokdad AH. A comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. Public Opin Q. 2008 Mar 20; 72(1):6–27.

18. Gaziano C. Comparative analysis of within-household respondent selection techniques. Public Opin Q. 2005 Mar 20; 69(1):124–57.

19. McCormick TH, Salganik MJ, Zheng T. How many people do you know?: efficiently estimating personal network size. J Am Stat Assoc. 2010; 105(489):59–70. PMID: 23729943

20. U. S. Bureau of the Census. State-to-state migration flows: 2013 [Internet]. 2013. Available: https:// www.census.gov/hhes/migration/data/acs/state-to-state.html.

21. U. S. Bureau of the Census. State-to-state migration flows: 2012 [Internet]. 2012. Available: https:// www.census.gov/hhes/migration/data/acs/state-to-state.html.

22. Kadushin C, Killworth PD, Bernard HR, Beveridge AA. Scale-up methods as applied to estimates of heroin use. J Drug Issues. 2006 Apr 1; 36(2):417–40.

23. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed-mode surveys: the tailored design method. Hoboken, NJ: John Wiley & Sons, Inc.; 2014.

24. Willis GB. Cognitive interviewing: a tool for improving questionnaire design. Thousand Oaks, CA US: SAGE Publications, Inc; 2005.