Social Science Research 62 (2017) 350–361

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

A comparison of network sampling designs for a hidden population of drug users: Random walk vs. respondent-driven sampling

David C. Bell [a, *], Elizabeth B. Erbaugh [b], Tabitha Serrano [c], Cheryl A. Dayton-Shotts [d], Isaac D. Montoya [1]

[a] Department of Sociology, Indiana University-Purdue University Indianapolis, 425 University Blvd, Indianapolis, IN 46202, USA
[b] Sociology and Anthropology, Stockton University, School of Social and Behavioral Sciences, 101 Vera King Farris Drive, Galloway, NJ 08205, USA
[c] Shell Oil Company, Houston, TX, USA
[d] University of Texas M.D. Anderson Cancer Center, USA

ARTICLE INFO

Article history:
Received 8 September 2014
Received in revised form 1 July 2016
Accepted 25 August 2016
Available online 27 August 2016

Keywords:
Network sampling
Hidden populations
Random walk
Respondent-driven sampling
Social networks

ABSTRACT

Both random walk and respondent-driven sampling (RDS) exploit social networks and may reduce biases introduced by earlier methods for sampling from hidden populations. Although RDS has become much more widely used by social researchers than random walk (RW), there has been little discussion of the tradeoffs in choosing RDS over RW. This paper compares experiences of implementing RW and RDS to recruit drug users to a network-based study in Houston, Texas. Both recruitment methods were implemented over comparable periods of time, with the same population, by the same research staff. RDS methods recruited more participants with less strain on staff. However, participants recruited through RW were more forthcoming than RDS participants in helping to recruit members of their social networks. Findings indicate that, dependent upon study goals, researchers' choice of design may influence participant recruitment, participant commitment, and impact on staff, factors that may in turn affect overall study success.

1. Introduction

Sampling designs such as outreach recruitment and targeted sampling have been used to recruit samples from hard-to-reach and hidden populations (Spreen, 1992). However, while these methods accomplish the goal of generating data, the external validity of the samples they generate may be limited by various biases. Newer, network-based sampling methods also allow the researcher to sample from a hidden population (Heckathorn, 1997; Klovdahl, 1985, 1989; Klovdahl et al., 1994; Spreen, 1992). Two designs, the random walk (Klovdahl, 1989, 1990; Liebow et al., 1995) and respondent-driven sampling (Broadhead et al., 1995; Heckathorn, 1997), take advantage of social networks within a population and aim to avoid some biases of earlier sampling methods. Appropriate analysis may allow the researcher to minimize biases associated with a given design in order to improve the estimate of population parameters (Gile and Handcock, 2011; Heckathorn, 2007; Thompson, 2011).

* Corresponding author.
E-mail addresses: dcbell@iupui.edu (D.C. Bell), erbaughe@stockton.edu (E.B. Erbaugh).
[1] Retired.

http://dx.doi.org/10.1016/j.ssresearch.2016.08.016
0049-089X/

Random walk (RW) sampling has been implemented in a relatively small number of studies. In contrast, the use of respondent-driven sampling (RDS) by social scientists has increased substantially in recent years (Gile and Handcock, 2010; Johnston and Sabin, 2010). The disparity in use of these sampling methods warrants comparative assessment of the two designs in practice. At present, little is known about the actual methods employed by members of a hidden population in the RDS recruiting process (Gile and Handcock, 2010). Short of having researchers accompany a sample of such recruiters in the field, the main means of studying differences between RW and RDS is to compare the responses and participation of those new members of the population who are recruited.

This paper compares experiences of using random walk and respondent-driven sampling to recruit participants from a high poverty, high drug use population. Prior studies have compared RDS to targeted sampling (Robinson et al., 2006; Rudolph et al., 2011) and to snowball sampling (Kendall et al., 2008); the current study is so far as we can determine the first to compare RW and RDS in side-by-side implementation. We identify tradeoffs in participant recruitment, participant commitment and staff impact between RW and RDS designs, factors that may in turn affect the representativeness of the sample, the quality of the data collected, and the overall success of the study.

1.1. Network-based sampling designs

1.1.1. Random walk sampling

The concept of interconnected personal and social networks is inherent to the random walk design (Klovdahl, 1989, 1990; Liebow et al., 1995). The random walk method was initiated as a link-tracing design in order to study structure in urban networks (Spreen, 1992), and was originally applied in neighborhood studies where persons were well and publicly known to one another (Sudman and Kalton, 1986; Sudman et al., 1988; van Meter, 1990). Nevertheless, some hidden populations, too, can be seen as collections of linked persons. A random walk can be conceptualized as a series of consecutive linkages from one person to another, and then to another (Sudman and Kalton, 1986; Sudman et al., 1988; van Meter, 1990). Each "step" in a random walk involves choosing a random member of the current participant's social network. Early motivations for using random walk with hidden populations included the ability to penetrate more deeply into the population from the initial sample, thereby achieving a more representative sample (Thompson, 2011).

One advantage of a random walk is that its procedures minimize *frame biases* (over- or under-representation of units or subgroups in creating the "list" of all elements in the target population). Random walks generate localized lists by soliciting the names of the peers and acquaintances of people from the target community (Klovdahl, 1985; McGrady et al., 1995). Staff recruiters generally use targeted sampling to select persons knowledgeable about the population as "seeds," each of whom is seen as connected directly and indirectly to other members of the population. Recruiters then randomly select names from lists of persons known by the "seed" individuals as targets for recruitment. As the random walk moves into the population, each person in the target population who is known by someone else in the population has a statistically non-zero chance of eventually being selected. Frame bias will increase if the population contains multiple networks that are not connected to one another. If all members of the population are connected (in what network researchers call a "connected component"), then all members are potentially reachable through one seed. Loners with no connections and members who belong to small components are liable to be excluded from the sampling frame. If a population contains multiple connected components, this bias can be reduced if the investigator selects multiple starting points ("seeds") in the different networks (Klovdahl, 1989, 1990; McGrady et al., 1995). Selecting multiple seeds aims to minimize sampling bias by finding various pathways into the social network. Random walks, if properly implemented, can thus yield a sample that is highly representative of the target population.

Another advantage of the random walk method is that *participation biases* (those resulting from individuals' unwillingness to participate, inability to participate, or incomplete participation) can be minimized. Recruitment success depends in part on the trust of the potential study participant in the recruiter. Such trust can be increased when the recruiter is introduced to the potential participant by a known member of the network, namely the informant whose list was used to select the potential participant (Sterk-Elifson, 1993; Sudman and Bradburn, 1982). Random walks thus have a built-in tendency to engender trust in participants.

However, a random walk can be expensive in terms of staff time and investment. *Implementation bias* (bias that may occur when researchers either avoid recruiting in certain areas or accept ineligible participants into a study) and/or *response bias* (the result of unusually high or low levels of openness, optimism, cooperation, attention or mood among participants) can be severe unless staff are able to cultivate participant trust and commitment (Sudman and Bradburn, 1982). Members of hidden populations, especially those whose members are engaged in illegal activities, are rarely eager to divulge personal information about themselves and their social networks to strangers (Liebow et al., 1995). Random walks may induce recruiters to be closely involved in the lives of their subjects because they must personally recruit each one, requiring extensive investment of time and resources (Klovdahl, 1990; McGrady et al., 1995).

1.1.2. Respondent-driven sampling

In respondent-driven sampling, the members of a hidden population themselves draw upon their own personal networks to recruit other members of the population (Heckathorn, 1997). Staff recruiters select seeds; seeds then become peer recruiters. Research staff tutor peer recruiters on study recruitment goals, give them a limited number of recruitment coupons, and usually offer them incentives for recruiting additional members of the target population. The peer recruiters then