

Using Respondent-Driven Sampling to Study Self-Employed Nigerians in New York City: Applicability for Immigrant Populations

Introduction

Sampling international immigrants is often a difficult task because international migrants constitute rare elements in the host population. They may also reject participation in research because of privacy concerns, regarding illegal residence status, or out of fear of outsiders or language barriers. They might not be residentially concentrated in ethnic neighborhoods and therefore difficult to locate. Also, for the majority of immigrant groups, no sampling frame exists so the size and boundaries of the population are unknown. Creating a sampling frame for these groups can become extremely costly. For these reasons, immigrants often form part of the so-called “hidden” populations.

Hidden populations have commonly been sampled using several methods, including target sampling, and snowball and other chain-referral techniques (Salganik and Heckathorn 2004). Target sampling involves sampling individuals in institutional settings (e.g. drug users in rehabilitation clinics, immigrants participating in associations.) This method can yield a large but biased sample, because individuals who enter institutional settings are a nonrandom section of the larger population.

Snowball sampling involves recruiting a few initial participants (seeds) for the study and they in turn recruit other people from their social networks, until the desired sample size is reached. The seeds can be selected in any manner, but are usually a convenience sample. The main advantage of this method is the ability to locate informants who otherwise would be difficult to include in the sample. Nonetheless, this method also has been severely criticized for its many flaws. First, the individuals recruited by the seeds are not randomly selected and therefore their characteristics are associated with those of their recruiter. Second, the sample is biased towards more cooperative individuals. Third, because referrals occur through network links, subjects with larger personal networks will be oversampled, and those largely isolated from the group may be missed (Salganik and Heckathorn 2004).

Recently, a new sampling technique called respondent-driven sampling (RDS) was introduced that enjoys the advantages of chain-referral techniques but, under specific guidelines, overcomes most of its limitations.

Respondent-Driven Sampling

RDS is conducted similarly to chain-referral (snowball) sampling techniques, but under certain conditions, it allows generalization of proportions from the sample to the population. That is, RDS converts snowball sampling techniques into probability sampling ones.

Unlike traditional probability sampling methods, RDS does not directly estimate from the sample to the population. Instead, it uses an indirect estimation method that first uses the sample to make estimates about the social network connecting the population. This information about the network is then used to derive the proportion of the population in different groups (Salganik and Heckathorn 2004).

RDS draws on Markov chains and the theory of biased networks to produce samples that are independent of the initial subjects from which sampling begins, therefore reducing potential bias introduced by seeds (Heckathorn 1997; Heckathorn 2002). Research has shown that if sufficient sampling waves are conducted, the final sample reaches equilibrium- that is, the sample is balanced with regards to a specific characteristic (such as ethnicity) regardless of who the seeds are. Exactly how many waves are necessary is currently being investigated, but RDS developers recommend at least six. Free software provided by the creator of RDS, Dr. Heckathorn, calculates when equilibrium has been reached in a particular sample.

Two important measures must be collected from every respondent. One is information about whom they were referred by. For this, a limited set of numbered coupons are given to each respondent, and they in turn give one to each referral (Heckathorn 1997). Second, the degree, or self-reported network size must be collected. There are several methods of collecting this information, including asking for the number of people known from certain categories (e.g. professors, doctors, Muslims), or the number of people from the target population known from certain spheres of life (family members, coworkers, church members) (Killworth, et al. 2003; McCarty, et al. 2001).

Several characteristics of the population's social networks can be inferred from sampling alone. Of particular interest to this project is the insularity index, which measures the degree to which respondents resemble one another with regards to any particular characteristic (Heckathorn 1997). For example, it can establish whether people tend to recruit others of their own gender for the project, or of their own ethnicity, neighborhood, etc. This information can be linked to the database created from the surveys to observe its relationship with a number of other variables, such as economic outcomes.

Respondent-driven sampling has been successfully used to study sex workers (Grazina Johnston, et al. 2006; Simic, et al. 2006), drug users (Heckathorn 1997; Robinson, et al. 2006) and jazz musicians (Heckathorn and Jeffri 2001). Its creator has noted that the method is relevant for the study of immigrant populations (Heckathorn 2006); this study undertook that challenge.

Sampling Nigerian Immigrants in New York City

Africans represent a small but quickly growing population of immigrants in the U.S. According to the 2000 U.S. Census, there were over one million African immigrants in the U.S., of whom almost 140,000 were Nigerian. This number is likely to have grown substantially since then, because almost 60% of all Africans in the 2000 Census arrived in the previous decade. New York City has the largest number of Nigerian residents and thus was chosen as the site for this study. Sampling Nigerians proved a difficult task because they are geographically dispersed all over the city and not concentrated in any particular neighborhood. The most concentrated Census tracts are less than 10% Nigerian.

Ethnographic fieldwork, begun early on in the study, eliminated other sampling strategies. Because the focus of the study are self-employed Nigerians, a large list of Nigerian businesses was compiled mainly from newspaper ads and referrals. This list excluded Nigerians who do not formally advertise their business, informal cab drivers, and non self-employed Nigerians also of interest in the survey. Sampling from this list was ruled out. Cluster sampling using Census tracts

as the primary units and sampling every Nigerian in the tract was also ruled out because it became evident that Nigerians were not necessarily aware of who their neighbors were beyond their block. RDS was selected as the sampling strategy because of its ability to sample hidden populations using respondents' social ties.

Seed Selection and Recruitment Process

Sampling began after some initial months of ethnographic fieldwork. Information from this period was incorporated in the selection of the five initial seeds to attempt to make them as diverse as possible and attempt to reduce bias from the beginning. In Nigeria, the three largest ethnic groups are the Hausa, Yoruba, and the Igbo. Nigerians in New York City seemed to be mainly Yoruba, numerous Igbo, and extremely few Hausa. Furthermore, various informants were from the minority Bini tribe.

Through qualitative interviews informants revealed that often their close friendships (besides their spouse) are with people of their own gender, and that they often belong to hometown associations, so high gender and ethnicity insularity was expected. Friendship ties between self-employed and wage laborers were common, so crossover between these employment categories was expected. In all, five seeds were selected with the following characteristics: a self-employed Bini man; a self-employed Yoruba man; a self-employed Yoruba woman; a self-employed Igbo woman; a Muslim man.

Seeds were given two coupons and asked if they could recruit another Nigerian immigrant living in New York City for the survey. Most chose to give the name and contact information of the people right there. A few called them on the cell phone to alert them that I would be contacting them. Others stated that they would first ask their friends for permission. One man said that as a teacher he understood the value of my project but that he would first explain to his friend the importance of helping me out before giving me their names.

Characteristics of the Sample

Sampling is still ongoing; therefore this abstract cannot include the results of the sampling process. All waves will be completed by March 2008. The following characteristics of the sample will be included and analyzed in the final paper:

Gender: We expect high gender insularity (respondents recruiting others of their same gender.) Whether men or women are linked is important to our project to understand how social networks affect the economic outcomes of men and women differently.

Ethnicity: RDS and its accompanying software have the capability of estimating group proportions of more than two groups, but more than four have never been measured. Nigerian immigrants belong to many ethnic groups. How RDS responds to so many categories will be assessed.

Self-employment status: We expect a low insularity index regarding self-employment status. This index can shed light into the factors that determine immigrant entrance into self-employment.

Religion: A high religion insularity index can determine the structure of immigrants' social networks and can help explain how these networks shape their economic opportunities.

In addition, a discussion section will be added where the advantages and disadvantages of RDS for immigrant populations is discussed. Nigerian immigrants differ from other immigrant populations in their ethnic diversity, high average educational level, and areas of economic insertion in the host society. Careful attention will be paid to discuss how the sampling technique might perform differently for varying immigrant populations.

Cited References

Grazina Johnston, Lisa , et al.

2006 Assessment of Respondent Driven Sampling for Recruiting Female Sex Workers in Two Vietnamese Cities: Reaching the Unseen Sex Worker. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 83(5):i16-i28.

Heckathorn, Douglas D.

1997 Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Forces* 44(2):174-199.

—

2002 Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Social Problems* 49(1):11-34.

—

2006 Studying Second-Generation Immigrants: Methodological Challenges and Innovative Solutions, Vol. 2007: Migration Information Source.

Heckathorn, Douglas D., and Joan Jeffri

2001 Finding the Beat: Using Respondent-Driven Sampling to Study Jazz Musicians *Poetics* 21:307-329.

Killworth, Peter D., et al.

2003 Two interpretations of reports of knowledge of subpopulation sizes. *Social Networks* 25:141-160.

McCarty, Christopher, et al.

2001 Comparing Two Methods for Estimating Network Size. *Human Organization* 60(1):28-39.

Robinson, William T., et al.

2006 Recruiting Injection Drug Users: A Three-Site Comparison of Results and Experiences with Respondent-Driven and Targeted Sampling Procedures. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 83(7):i29-i38.

Salganik, Matthew J., and Douglas D. Heckathorn

2004 Sampling and Estimation in Hidden Populations using Respondent-Driven Sampling. *Sociological Methodology*:193-239.

Simic, Milena, et al.

2006 Exploring Barriers to "Respondent Driven Sampling" in Sex Worker and Drug-Injecting Sex Worker Populations in Eastern Europe. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 83(7):i6-i15