Overview of the Special Issue on Surveying the Hard-to-Reach

Gordon B. Willis[1], Tom W. Smith[2], Salma Shariff-Marco[3], and Ned English[4]

1. Introduction

Throughout the course of the development of survey methods, critical concerns have arisen in tandem with changes in society that impact the nature and composition of the populations that researchers endeavor to understand. In recent years, a concern that has attracted considerable methodological interest involves the conducting of surveys that include members of so-called Hard-to-Reach (H2R) groups. H2R groups have become increasingly important to include within a range of population surveys, given both a burgeoning emphasis on representation of demographic subgroups (e.g., Asians within the U.S. population), and on groups that are of interest due to their potentially unique characteristics or sociocultural location (e.g., transgender individuals).

To further methodological work in this area, the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations was held from October 31 to November 3, 2012 in New Orleans. This gathering aimed to connect researchers across a range of fields – including survey methodology, statistics, demography, sociology, anthropology, ethnography, and psychology – and engage them in discussions devoted to advancing our methodology for surveying groups that have proven difficult to include in population surveys. However, from the start this seemingly targeted and even niche-like area of interest proved somewhat hard to encapsulate in a simple definition, and finding solutions turned out to be even more challenging. In particular, the science of 'reaching the H2R' requires that we address two vital issues: (a) who in particular are we talking about when we use the label 'hard-to-reach'? and (b) what do we mean by 'reaching' them? Considering both of these challenges, the conference advertisement (currently available at http://www.amstat.org/meetings/h2r/2012/pdfs/H2R2012Flyer.pdf) lists a range of topics and subpopulations of interest, and includes the following as examples of H2R groups:

(a) Racial minorities
(b) Immigrant populations
(c) Indigenous populations

[1] Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institute of Health, 9609 Medical Center Drive, Bethesda, Maryland, 20892-9762, U.S.A. Email: willisg@mail.nih.gov
[2] NORC, University of Chicago, 1155 East 60th street, Chicago, IL 60637, U.S.A. Email: smith-tom@norc.org
[3] Cancer Prevention Institute of California, 2201 Walnut Ave, Suite 300, Fremont, CA 94538, U.S.A. Email: salma.shariff-marco@cpic.org
[4] NORC at the University of Chicago, Statistics and Methodology, 55 E. Monroe St., Suite 2000, Chicago, IL 60647, U.S.A. Email: ENGLISH-NED@norc.org

(d) Highly mobile and migrant populations
(e) Homeless and refugee populations
(f) Sexual minorities
(g) Populations affected by natural disasters
(h) Populations in zones of armed conflict
(i) Stigmatized populations
(j) Cross-cultural similarities and differences in H2R populations
(k) Linguistic and cultural minorities

Concerning the second element defining H2R groups, or what we mean when we refer to 'reaching' them, Tourangeau (2014) has provided a helpful model within the opening chapter of a book deriving from the conference (Tourangeau et al. 2014). Tourangeau's model delineates various points in the survey process that present difficulty, as categorized into the following bins:

(1) Sampling/Coverage: Those who are difficult to include in a statistical sample;
(2) Identification: Those who are difficult to identify as eligible survey respondents;
(3) Location/Contact: Those who are difficult to find and to make contact with for purposes of engaging in a survey;
(4) Persuasion: Those who are difficult to convince to take part in a survey, once located;
(5) Interviewing: Those who are willing to be interviewed, but who are difficult to successfully interview.

Survey methodologists will recognize the close association between these concepts and the related statistical error subtypes of coverage, sampling, nonresponse, and response error.

2. The JOS Special Issue

The articles in this volume derive from contributed papers delivered at the conference, and supplement those commissioned for a book that includes chapters representing conference invited papers (Tourangeau et al. 2014). These special issue contributions address varying facets of the hard-to-reach continuum described by Tourangeau. For purposes of simplification – and perhaps based on the notion that "good things come in threes" – we have further aggregated these five elements into three general factors, each representing a basic challenge to successfully interviewing someone we think of as hard-to-reach:

(1) *Selection*: Choosing *who* we are attempting to interview (or to 'enumerate,' in the case that we are counting as opposed to collecting self-report information). Sampling and coverage issues dominate, along with challenges of identification.
(2) *Recruitment*: Deciding *how to locate* potential respondents, and, in cases where self-report is required, *how to persuade* them to consent to participating in the survey, once they have been sampled and identified.
(3) *Interviewing*: Determining *how to conduct* the interview. Beyond sampling, identifying, locating, and persuading a potential respondent, we must also consider the survey administration mode to be used (e.g., internet, telephone, mail), as well as

Table 1.   Journal of Official Statistics, Special Issue on Surveying the Hard to Reach: Attention devoted by each of the ten articles to three basics challenges in surveying the H2R.

Special Issue article:	(1) Selection Who to interview: Sampling and Identifying	(2) Recruitment How to 'get' the respondent: Locating, Contacting, Persuading	(3) Interviewing How to complete the interview: Administration mode, interviewer, etc.
1) Potential Uses of Administrative Records for Triple System Modeling for Estimation of Census Coverage Error in 2020: Griffin	X		
2) Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations: Himelein, Eckman, and Murray	X		
3) Enumerating the Hidden Homeless: Strategies to Estimate the Homeless Gone Missing From a Point-in-Time Count. Agans, Jefferson, Bowling, Zeng, Yang, and Silverbush	X		
4) A Study of Assimilation Bias in Name-Based Sampling of Migrants: Schnell, Trappmann, and Gramlich	X		
5) Comparing Survey and Sampling Methods for Reaching Sexual Minority Individuals in Flanders: Dewaele, Caen, and Buysse	X	X	
6) A City-Based Design That Attempts to Improve National Representativeness of Asians: Pedlow	X	X	
7) Recruiting an Internet Panel Using Respondent-Driven Sampling: Schonlau, Weidmer, and Kapteyn	X	X	
8) Locating Longitudinal Respondents After a 50-Year Hiatus: Stone, Scott, Battle, and Maher		X	
9) Evaluating the Efficiency of Methods to Recruit Asian Research Participants: Park and Sha		X	
10) Reaching Hard-to-Survey Populations: Mode Choice and Mode Preference: Haan, Ongena, and Aarts			X

who should conduct the interview, and what other procedural parameters may be optimal for obtaining truthful and accurate responses.

In the interest of organizational clarity, we categorize the articles in the current volume according to these three factors, as illustrated in Table 1. The table depicts what we consider to be the main elements addressed by each author's contribution to the special issue (with the caveat that each research effort may span multiple areas, and our assignment to category may be arguable due to the lack of firm boundaries between them).

Articles involving selection. It is clear that across these articles, significant attention is paid to sampling and coverage issues – that is, with respect to *constructing* the sample frame; seven of the ten articles attend to this issue. Sample frame construction is vital for addressing the companion issue of coverage – ensuring that members of the desired population are adequately represented in the sampling frame. For study of H2R populations, a major challenge to coverage is the needle-in-a-haystack phenomenon, especially where the required population 'units' (i.e., people) are well-hidden among a larger, general population – for example, the homeless, or migrants. First, the article by Griffin focuses on the measurement of coverage error in the U.S. Census in a manner relevant to H2R subpopulations, through reference to administrative records. Himelein, Eckman, and Murray tackle the vexing challenge of surveying a special type of group – nomads – who have no set residential location, and who therefore depart from our usual notions of 'place of residence.' Similarly, Agans, Jefferson, Bowling, Zeng, Yang, and Silverbush consider the way in which members of household units can be relied upon to identify individuals – the homeless – who also have no set place of residence but who may have existing relationships with those who do (i.e., individuals who provide temporary shelter).

Several other articles focus on populations whose members do have a set place of residence, but who are difficult for the survey takers to identify and enumerate as members of an H2R group because they may be 'hidden in plain sight.' Schnell, Trappman, and Gramlich describe a study that involves the use of name-based sampling to create a frame of immigrants contained within society at large. Dewaele, Caen, and Buysse also focus on an H2R population – sexual minorities – who are integrated within the larger population, but who are not readily identifiable with respect to H2R status through any means other than self-identification. On the other hand, research by Pedlow attempts to leverage the fact that some subpopulations (in this case, Asians) do tend to be physically clustered, and can be sampled via a geographically oriented (city-based) sampling approach. Finally, Schonlau, Weidmer, and Kapteyn investigate the development of a sample frame through the use of *respondent-driven sampling,* in which no suitable sampling frame exists or can reasonably be constructed by the researchers, and instead relies upon respondents from a particular H2R group to themselves produce contact information for additional eligible individuals.

Articles involving recruitment. The second major factor that we have defined involves recruitment, which can be viewed as literally 'reaching' the respondents, and two of the manuscripts focus mainly on this step. Stone, Scott, Battle, and Mahar regard H2R status as imposed by the calendar. Although the sought-after respondents within this survey were not demographically unique, a fifty-year follow-up interval rendered them literally hard to

reach, and put a premium on methodology related to respondent tracing. Once the identified individuals are located through appropriate detective work, the surveyor then must begin the process of selling the survey, persuading the located individual to participate. Park and Sha consider recruitment from a somewhat different vantage point, concerning how to locate and persuade monolingual Asians for purposes of pretesting survey questionnaires. The obvious solution is to use the language that respondents (literally) use; what is less obvious, and is intensively investigated by the authors by relying on multiple efficiency metrics, is the medium to be employed: flyers, newspaper advertisements, word-of-mouth, or something else that may be specific to the H2R population.

Articles involving interviewing. A final article, by Haan, Ongena, and Aarts, mainly addresses the conduct of the survey interview, or how to reach respondents in terms of presenting and then obtaining information in a way that makes sense to them. The key consideration is choice of administration mode, which involves factors related both to access (e.g., do potential respondents have internet service) and social dynamics (do they prefer interaction with a human interviewer – that is, interviewer administration, or would they rather answer a computer screen or a piece of paper, under questionnaire self-administration?). The dynamics of interviewing could address a range of other parameters as well, such as interviewer demographic characteristics and behavior, or the physical location of the interview (at home, or elsewhere).

3. Looking to the Future

The articles in this volume attempt to provide answers to the issues listed above, in most cases through use of examples and case studies. However, they also raise fundamental issues that challenge the fledgling science of surveying the hard-to-reach. Reflection on the difficulties and barriers to the enumeration process, or to the conduct of self-report surveys, may even suggest the need for a subtle but important shift in investigator viewpoint – and perhaps in nomenclature. We must recognize that, from the respondent's point of view, he or she may not be at all 'hard to reach.' There are certainly subtypes of potential respondents who truly are challenging to select, recruit, and interview because those individuals take steps to make each of these steps difficult (e.g., undocumented individuals who hide from the administrators of a Federal survey). There are others, however, who present difficulties mainly from the perspective of the researcher. We may fail to reach monolingual Vietnamese speakers largely because we fail to enlist interviewers who can communicate in that language. By way of analogy, one can state that Timbuktu is hard to reach – but this is true only from the vantage point of Western locations, and not if one begins the trip from a nearby town in Mali. As such, the application of the label 'H2R' to a particular group may mainly reflect the separation between researcher and respondent, rather than some immutable characteristic of the latter.

An alternative to 'Hard to Reach' is suggested by Tourangeau (2014), who has selected the general term 'Hard to Survey,' and who regards 'Hard-to-Reach' as one subcategory of specific difficulties encountered (literally, those who are hard to locate and to contact, once they have been identified). This solution may not placate those who object that the general use of the term 'Hard' carries the implication that such members of some groups are 'resistant' or 'uncooperative.' As an alternative, a judgment-neutral approach might be

to state simply that some groups are 'historically under-represented' to convey the notion that certain populations have tended to be left out of survey (and other) research.

A final, opposing perspective is that survey methodologists need not be overly concerned with labeling, but rather with the ultimate outcome of their survey practices. If researchers increasingly are committed to enhancing our capacities for including those who are left out, and are committed to expending the appropriate resources and/or effort (as are those represented in the current volume), then it may not matter whether they regard their respondents as hard-to-reach, as under-represented, or – from the point of view of staff on the ground – as 'difficult completions.' It is our hope and intent that the directions defined within the current set of manuscripts provide a path towards the continued development of imaginative and effective methodologies for ensuring that our surveys and Censuses are equally inclusive of all.

4.   References

Tourangeau, R. (2014). Defining Hard-to-Survey Populations. In Hard-to-Survey Populations, R. Tourangeau, B. Edwards, T.P. Johnson, K.M. Wolter, and N. Bates (eds). Cambridge: Cambridge University Press, (In Press).

Tourangeau, R., Edwards, B., Johnson, T.P., Wolter, K.M., and Bates, N. (2014). Hard-to-Survey Populations. Cambridge: Cambridge University Press, (In Press).