



Defining and Measuring Hard-to-Survey Populations

2014 FCSM Statistical Policy Seminar

Roger Tourangeau, Westat

December 15, 2014

Defining Hard-to-Survey Populations

- In an era of falling response rates, it may seem like they are all hard-to-survey
- Still, some are harder than others
- What are the main sources of trouble? It is *not* just “rare and elusive” populations

Defining Hard-to-Survey Populations--II

- Hard-to-survey populations are those that create problems for one or more key survey operations
 - Hard-to-sample (rare populations with no population-specific frame)
 - Hard-to-identify (based on hidden or stigmatizing characteristics)
 - Hard-to-find or contact (mobile populations; pop's with access issues)
 - Hard-to-persuade (resistant; alienated)
 - Hard-to-interview (language barriers)

Hard-to-Sample Populations

- Much of the sampling literature on H2S populations focuses on the issue of rarity
 - How do you sample efficiently when the population is small and must be found by screening members of a larger population?
 - E.g., NLSY-97 looking for young people (ages 12-23)
- Several factors determine the level of difficulty
 - Rarity or prevalence (Rare population as a fraction of the larger population covered on the frame)
 - One thing that helps is concentration in a few high density strata or areas (only works if most of the rare population in these areas)
 - Relative cost of screening vs. main interview

Hard-to-Sample Populations: Metrics

- Simple index is relative increase in field costs due to the need for screening

$$\Delta_c = 1 + c / P,$$

where P is the overall prevalence of the rare population and c is the cost of a screener relative to the main interview

- For example, if the rare population is .05 of the general population and the screening cost is .10 of the main interview cost, then the screening costs triple the field costs

Hard-to-Identify

- When you have to screen, the assumption is that the screener informant both knows the relevant characteristics and is willing to report them
- Motivated misreporting: Deliberate concealment by the R or “steering” by the I
- Tourangeau, Kearney, Shapiro, and Ernst (1996): In poor neighborhoods, anonymous rostering found 30% more young black males; screener R’s worried about negative consequences of listing them
- Tourangeau, Kreuter, and Eckman (2012): Full roster found more members of the target population than direct question (“Is anyone here from 35 to 55 years old?”)—45% versus 32%

Hard-to-Locate, Hard-to-find

- Two issues here:
 - Can you find the members of the population?
 - Can you reach them?
- Hard-to-locate: Often a big problem here is mobility, or more generally, people not linked to a specific residence
 - Members of traditionally nomadic cultures (such as the Bedouins of Southwest Asia and the Tuareg of North Africa);
 - Itinerant minorities (such as the Romani in Europe or the Travellers in Ireland);
 - Persons who are temporarily mobile or displaced (recent immigrants, homeless persons, refugees); and
 - Persons at a mobile stage in their life cycle (young people).

Hard-to-Locate—II

- Tracking an issue with longitudinal surveys (Couper and Ofstedahl, 2009; Lepkowski and Couper, 2002)
- Couper and Ofstedahl look at PSID and HRS
- 13.7 percent of the US population moved in 2004; lower rates in Western Europe
- PSID located 96.7 percent of the 1,441 cases that needed to be tracked for the 2003 round and the HRS located 98.7 percent of its 1,294 movers for the 2004 round
- Successful but a drain on resources (10.2 tracking calls for PSID, 7.4 for HRS)

Hard-to-Contact

- With groups, the issue is not finding them but getting to them
 - Gated communities, locked apartment buildings
 - Nursery homes
 - Other populations with access issues (children)
 - Cell phones—Do they make things better or worse?
- Metrics
 - Contact rate (related notion: Average contact probability)
 - Number of calls, contact attempts per case, average before contact made
 - Lepkowski and Couper (2002) advocate a two stage model—
probability of finding \times probability of making contact

Hard-to-Persuade

- Some groups seem chronically harder to get to participate than others
 - Younger people (versus older people)
 - Men (versus women)
 - People living in big cities (versus the suburbs and rural areas)
 - The less-educated (versus the well educated)
 - The poor (versus the well-to-do)

Hard-to-Persuade—II

- Two general hypotheses about who takes part readily
 - Community involvement (Abraham, Bianchi, and Maitland, 2006); civic engagement (Groves, Singer, and Corning, 2000); volunteering (Abraham, Helms, and Presser (2009); voting (Tourangeau, Groves, and Redline, 2010)
 - Busyness
- Salience-leverage theory suggests that more transitory factors are more important in decision to participate; features of survey that happen to be salient
 - Topic interest (“Birder” study—Groves et al., 2006)
 - “Avidity bias” recreationists more likely to take part in a survey involving their activity

Hard-to-Persuade—Metrics

- Metrics
 - The refusal rate (the proportion of those who were contacted but who declined to take part)
 - The proportion of sample members who required refusal conversion, special incentives, or other extraordinary measures to obtain their cooperation.

Hard-to-Interview

- Three other reasons why some pop's are difficult to survey:
 - They may be vulnerable populations (prisoners, young children), requiring explicit consent;
 - They may have impairments that makes them difficult to interview at least under the standard survey protocols; or
 - They may not speak (or read) the language in which the survey questionnaire is written.
- Surveys generally designed for people in reasonably good health, with intellectual abilities in the normal range (or above), and without serious sensory impairments. Those outside this range left out of many surveys.
- These exclusions often not a problem for surveys of the general population, since these conditions are rare. These sources of difficulty (including language barriers) account for less than eight percent of CPS nonresponse (vs. 53 percent for refusal)

Hard-to-Interview—II

- Surveys do various things to cope
 - Translate the q'aire into multiple languages
 - TTY (text telephone) for the deaf
 - ECLS-K: Cognitive assessments for non-readers
- Proxies
- Metrics
 - Proportion unable to do the interview
 - Proportion requiring a proxy
 - Added cost per case (Δ_c) due to special accommodations

General Metrics for Difficulty

- U.S. and U.K. have indices for hard-to-count (HtC) areas for censuses
- U.S. index based on 12 variables (Robinson, Johanson, and Bruce, 2007)
 - Percent of dwelling units that are vacant;
 - Percent that are not single-family units;
 - Percent of occupied units that are occupied by renters;
 - Percent of occupied units with more than 1.5 persons per room;
 - Percent of households that are not husband/wife families;
 - Percent of occupied units with no telephone service;
 - Percent of persons below the poverty line;
 - Percent of households getting public assistance;
 - Percent of persons over 16 who are unemployed;
 - Percent of households where none of the adults (over 14) speak English well;
 - Percent of households that moved in the past year;

General Metrics for Difficulty—II

- Each CT gets a score on each dimension (0 to 11)
- CT level scores; correlate .77 with return rate in Census 2000
- Mix of variables:
 - SES, poverty
 - Mobility
 - Structure-related variables (apartments vs. single-family homes)
 - Language variables
- Similar mix of variables in UK index (Compton and Abbott, 2012)

General Metrics for Difficulty—III

- UK HtC Index (Compton and Abbott)
- Group areas with ~750 households based on five variables:
 - The proportion of people claiming Income Support or Jobseeker's Allowance;
 - The proportion of young people;
 - The proportion of people who are not 'White British';
 - The relative house price within a local authority; and
 - The density of dwellings in an area.
- Fit logistic regression model based on these variables
- Group areas into five strata based on predicted response rate

Summary

- Hard-to-survey populations can be high bias or high cost populations
- Why high bias?
 - Hard-to-sample populations may be undercovered or may have to be sampled via non-probability methods
 - Hard-to-identify may be underrepresented (for example, due to screening errors)
 - Hard-to-locate or contact may have high rates of nonresponse
 - Hard-to-persuade may have high refusal rates, leading to nr bias
 - Hard-to-interview may be underrepresented due to nonresponse
- Why high cost?
 - Added screening costs
 - Added costs to contact and locate, etc

Thank You!!!