Modernisation of Statistics Production (MSP2009)
2-4 November 2009

Session: Management of Processes and Resources

Adapting to the Challenges of Household Survey Data Collection in a Multimode Environment

Colm O'Muircheartaigh
Dean and Professor, Harris School of Public Policy, and Senior Fellow, National Opinion Research Center
University of Chicago
colm@uchicago.edu

Each of the major modes of survey data collection faces significant challenges. For face-to-face surveys, costs of travel and interviewing have increased exponentially. For telephone surveys, the coverage of telephone frames has declined and continues to decline, especially for younger people, while the response rates to telephone approaches is plummeting. For mail surveys, respondent selection and response rates continue to present problems. For internet surveys, no satisfactory frame yet exists for general population surveys.

Four other developments contribute to the changing survey landscape. Surveys are being used increasingly to collect biomeasures as well as reports of behavior and attitudes. Databases with high levels of coverage of the household population continue to be developed and are becoming more accessible to researchers. The flexibility of IT systems is increasing, and nimble manipulation of cases is becoming possible. Certain kinds of data about individuals are becoming more readily available to the individuals themselves, and may be transferable to others with (relatively) little burden on survey respondents.

The paper considers the strategic design of surveys in this context. An illustration from the *REACH U*S evaluation surveys [a set of health behavior surveys in minority communities] is presented.

Key words: Survey design; address based sampling; ABS; M-ABS

1. Exploiting Diversity in Populations and Modes: the Multimode Environment for Survey Research

This paper addresses data collection for sample surveys. I will confine myself to a discussion of data collection for household surveys, and the examples will come primarily from surveys in the USA. Many of the issues generalize however to surveys of establishments and to the general situation in many countries.

From the introduction of the sample survey as a methodology for data collection for human populations in the 1890s (Kiaer, 1897), surveys followed similar protocols as those used for the complete enumerations they were designed to replace. The dominance of probability sampling that followed Neyman's comparison of random and purposive sampling (Neyman, 1934) established the scientific foundation for sample design.

Standards for questionnaire design were developed through the 1940s, and interviewer training methods were established by the mid 1940s. By the 1950s a homogeneous model for the social survey had emerged. Sample designs were standardized in most countries, only one mode of data collection was used for each survey, questionnaires were standardized and monolingual, and generally dealt with a single unit of measurement, either an individual or a household.

1.1 Problems with Current Surveys

For almost half a century, the standard model for the sample survey functioned effectively and survived basically intact. In the 1960s and 1970s, telephone surveys were introduced on a wide scale, first in market research and then in government work. Sampling frames were developed, question design issues were addressed, and response rates were respectably high. The portable computer led to development of computer-assisted survey information collection (CASIC: CATI for telephone surveys, CAPI for in person surveys, CASI for self-administered questionnaires, and CADE for third-party data entry).These were incorporated into the process without any fundamental disruption of the underlying model.

In recent years, there has been a gradual deterioration in the quality of surveys. Though some surveys still maintain very high standards, this has been achieved through a substantial increase in the effort and expense. Face-to-face surveys are generally too slow and too expensive, and even with the added time and expense response rates are difficult to maintain at previous levels. Telephone surveys suffer from coverage problems (especially with cell-phones, or if a list frame is used), density problems (if a comprehensive frame of possible telephone numbers is used), and low response rates. Response rates have declined steadily since telephone surveys became widespread in the 1970s, and now hover near 30% for high quality surveys (much lower in many circumstances);. Internet surveys suffer from the absence of any population frame that can generate a probability sample of the population. Other personal electronic devices have too little penetration, lack frames, and their behavior is not well understood. Mail surveys have problems with response rates and within-household respondent selection; no satisfactory way has yet been found to deal with subsampling within households by mail.

As a response to the gradual deterioration of the standard model, the social survey has undergone a transformation in recent years. Diversity has been recognized and exploited; technology has been harnessed to the service of surveys, the range of measures surveys can collect has expanded, and modes thought to be obsolete have been resurrected.

1.2 Recognizing Diversity: An Illustration

An example may illustrate one way in which diversity has been recognized. The standard sample design for face-to-face surveys in the USA was a multi-stage area sample using metropolitan areas or counties as primary sampling units, with the selection of either one or two further area stages using smaller compact areas, and with field listing of household units within quite small ultimate area units. In the past ten years or so, three developments have allowed samplers to reconsider their designs. First, a high quality list of addresses (based on the US Postal Service’s Delivery Sequence File (address list)) has

become available commercially. Second, geocoding software has made it possible to locate these addresses in Census blocks and other spatial units. Third, a pre-fieldwork assessment of the quality of these addresses is possible. Consequently the US can be treated not as an undifferentiated whole, but as a set of strata with quite different design implications.

Though the USA is correctly seen as a large sparsely populated country, when we consider the variability in population density across different geographical areas it may more accurately be considered as a combination/mixture of populations of quite different character. Figure 1 shows the most densely populated Standard Metropolitan Statistical Areas in the USA. The densely colored core represents the parts of these SMSAs in which dwellings have street-style addresses (and thus for which we have a list frame). These SMSAs are sufficiently large that they are certainty selections in most national sample designs.

Figure 1: High-Density Urban

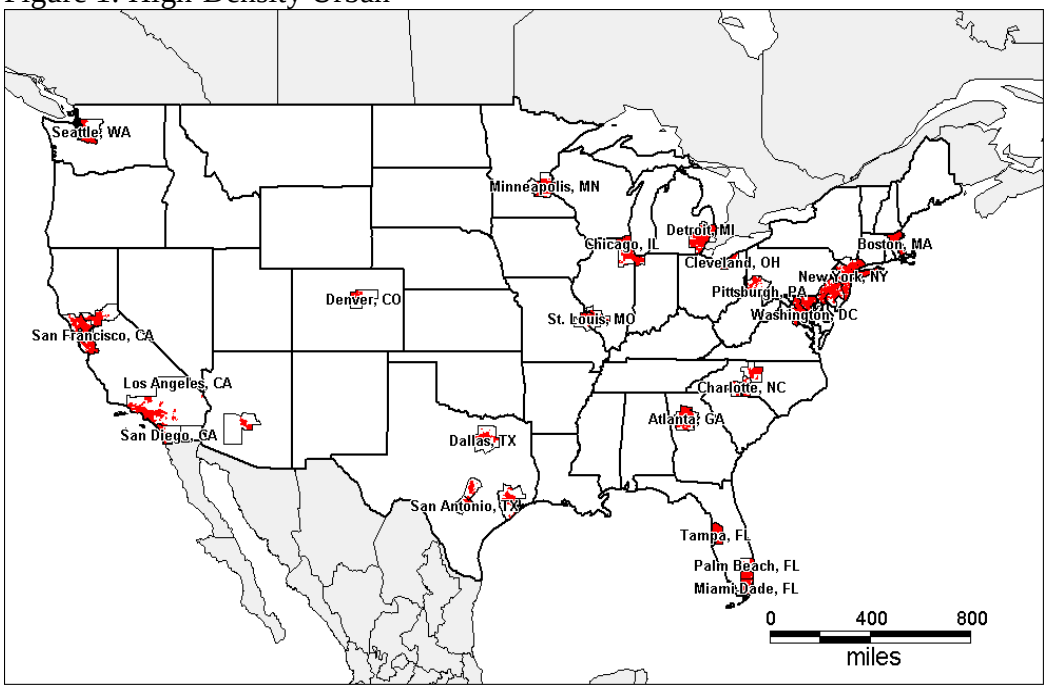


Figure 2 shows the next level of density and the parts of these SMSAs which have street-style addresses. For cost reasons a sample of these less dense (and smaller in population) areas must be selected in the first stage of the design. Within the selected areas, however, once again a list frame is available, making the choice of secondary sampling unit relatively invariant in terms of cost.

Finally in the rest of the country (Figure 3) the situation resembles more the standard area probability sampling situation, in which lists must be constructed after selection of area units. This constrains the size of these units due to the considerable cost of list construction.

Figure 2: Other Major Urban and Suburban

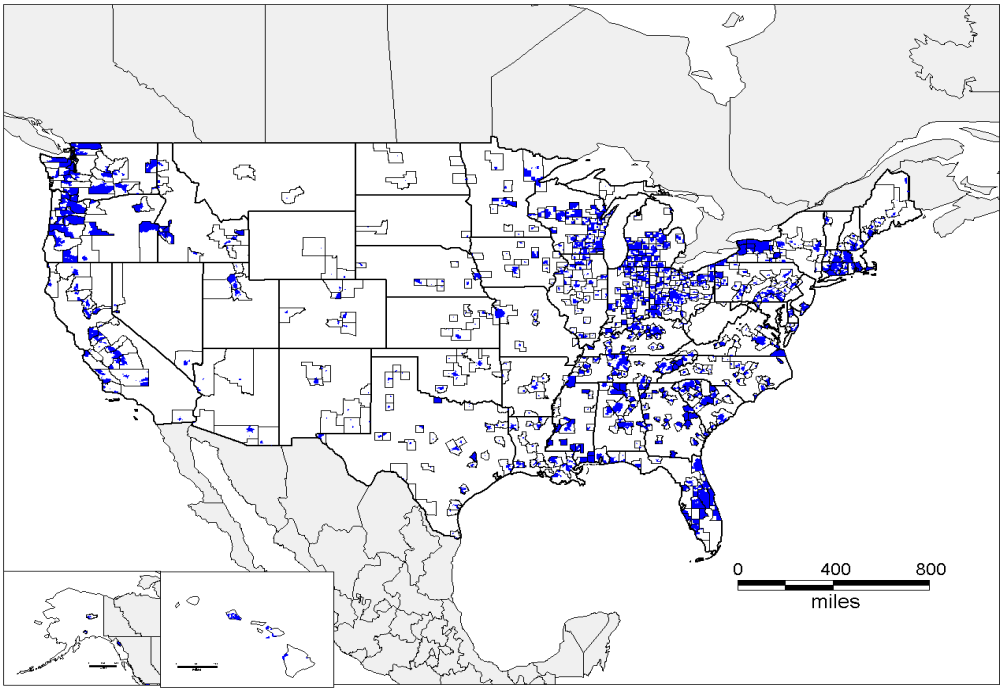


Figure 3: Small Counties and Non-Urban

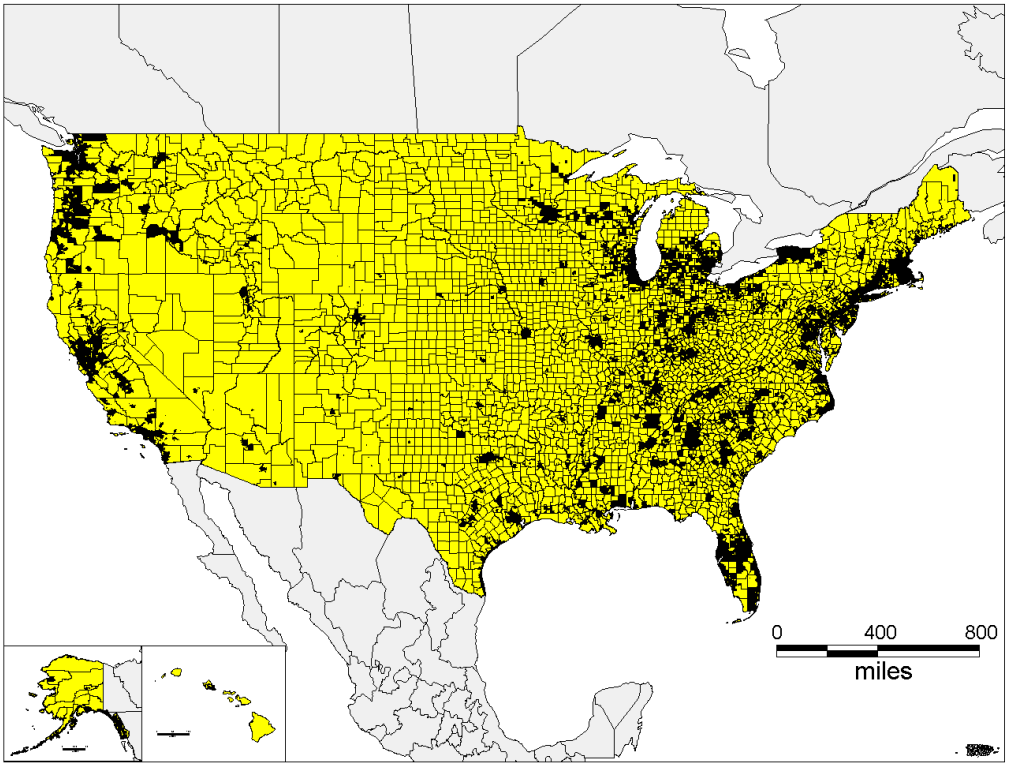
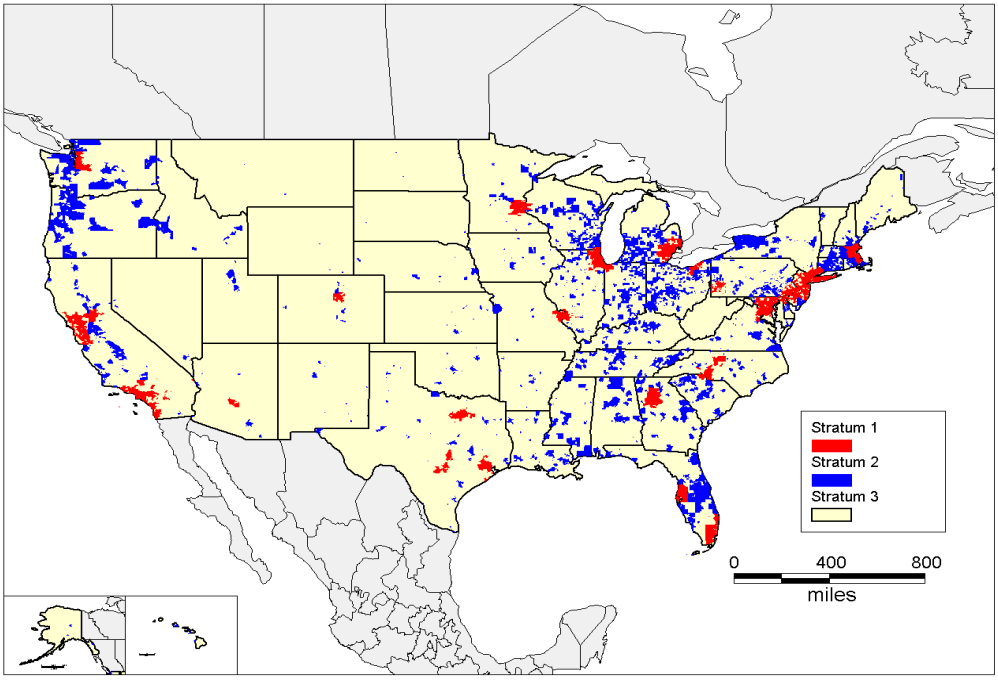


Figure 4 shows the distribution of the three strata across the US.



Stratum 1 has 41% of the population in only 2% of the area of the country. This population density is comparable to that of densely populated areas in northern Europe and elsewhere. Stratum 2 has 32% of the population in a further 6% of the area – a population density somewhere between inner urban and suburban. The remaining 92% of the area contains only a quarter (about 27%) of the population. It is only in this stratum (stratum 3) that tight clustering of households is necessary at the final stage of selection.

Thus each stratum has quite different design (cost/variance) parameters; this should be reflected in the ultimate design.

2. Changes in Survey Environment

2.1 Biomeasures

Historically, biological data collection was included in some survey work, but the collection was carried out by medically qualified personnel, with the recruitment (and perhaps questionnaire data) being the responsibility of the survey interviewers. In the past 10 years there has been a dramatic increase in the incorporation of biomeasure collection directly in the work of the interviewers. This has become possible mainly because of the extraordinary advances in biological assay of samples that are minimally invasive, such as blood spots (obtained from finger pricks) and saliva. In the US at least, respondents have also shown themselves to be comfortable with requests for more intimate data such as self-administered vaginal swabs. This extension of the scope of surveys also has implications for survey design.

2.2 Technology

Advances in technology and its application have also changed the options available for survey design and implementation in three (TK) ways. First, computer-assisted sampling is much faster than traditional list- and map-based selection and it is in principle portable across modes and frames. Second, Geographical Information Systems (GIS) can be used to link frames generated from information previously confined to different modes. Third, Computer-assisted Survey Information Collection (CASIC) has been adapted to multiple modes; can (in principle) provide flexible and easily controlled delivery of a tailored stimulus; can record complex (bio)measurements directly; can record location and movement; and can send real-time data to survey headquarters.

3. The New Approach – Matching Diversity with Diversity

In order to overcome the varied problems presented by the different approaches, we need to re-think our approach to data collection. Consider the problem. The world/population, our target characteristics, and the survey objectives vary in a multitude of ways.

First, we have multiple frames that purport to describe the population – most are specific to a particular mode of selection (telephone numbers and mailing addresses, for instance). The element in each frame may be different: landlines typically identify households or dwelling units (as do addresses), cell-phones typically relate to individuals.

Second, the characteristics we seek to measure may make very different demands on the data collection process. Observational data require a visit from the interviewer. Biomeasures and other physical measurements typically require the presence of the interviewer and the respondent in the same place. Attitude data may be collected by any of a number of modes, though there may be arguments as to the relative quality of the data.

Third, in the data collection process there are multiple tasks, and the information and skills required for each may be quite different. Thus locating and identifying the selected sample member may be a task that requires technological skills and care; convincing the selected unit to participate in the survey requires sales and negotiation skills that involve building a personal connection with the potential respondent; and administering the interview schedule correctly requires care and adherence to instructions (and perhaps some (psychological) distance between the interviewer and the respondent.

Fourth, the platform on which the data collection is built may also vary. There may be a paper instrument, and the electronic version may be on the interviewer's laptop computer, in a form suitable for interviews or self-completion, or perhaps on the web.

Fifth, the accessibility of the population may also be highly variable. The differences are stark even by looking at the range across the three strata illustrated in figures 1 through 3. Even taking the average of these densities, population density in stratum 1 is 70 times that of stratum 3; the ratio of the extremes in these strata is of course much higher; for

example, the ratio of the density of the top half of stratum 1 to the density of the bottom half of stratum 3 is 200.

Though the variations described above occur within the US population, their implications become even clearer when trying to design studies across countries and across cultures (multicultural and multinational surveys). The futility of trying to standardize sample designs across countries has long been recognized (see, for instance, Verma, Scott, and O'Muircheartaigh (1980) for the World Fertility Survey); there has been more resistance to the idea of standardizing objectives rather than process in other areas of design.

3.1 Some Dimensions of Variation

To illustrate the complexity of the situation, consider the characteristics described briefly in the table below.

Aspect	**Variation across…**
Nature and quality of the frame	Home address, phone number, internet address
Capacity to respond	Health, education, age, ...
Propensity to respond	Culture, personality, …
Population diversity	Language, cognitive capacity, robustness to burden
Particular mode-specific Considerations	e.g. coverage for internet, population density for face-to-face, …

3.2 Some Cells of the Multidimensional Lattice

Two examples illustrate the extremes for a hypothetical survey of individuals in the US.

Ideal: On the street-style address frame; living in high-density population area; listed landline phone number; available cellphone number and individual-specific internet (frame-linked) address; regular internet access; high propensity to respond to all approaches. Such an individual is available through multiple frames, is willing to provide data through multiple modes; and provides high quality data through all approaches. Our ideal prospect.

Not good: Not on the street-style address frame; living in low-density population area; no phone; no internet access; low propensity to respond to any and all approaches. Available only by expensive field listing and face to face interviewing, a poor prospect for survey research.

3.3 Multiple Modes, Multiple Sub-Designs

With such a population, it seems unreasonable to design a survey with only one frame and one mode. The complex lattice of the population characteristics should be matched by a coherent lattice design. The design should make it posssible to obtain data from the ideal elements cheaply and efficiently, and save the more complex and expensive procedures for the more difficult cases.

Easier said than done, however, as identifying the characteristics of the cases in advance is well-nigh impossible (though some of the characteristics (population density, quality of address) may be known. The characteristics that come to light only during the process

(propensity to respond to a particular mode, for instance) present a different problem. The process must be sufficiently flexible and agile to make dynamic adjustments during the data collection. At a macro level the survey process must produce real time streams of information on dispositions of cases, and be able to produce projections of outcomes so that the issued sample can be adjusted depending on progress [this is particularly the case for situations in which the frame is seriously imperfect so that eligibility rates are not known in advance. It must be possible to switch fluently among modes so that should it become apparent that a respondent is more comfortable with a particular mode that mode can be offered. It would also be desirable to be able to switch among burdens (by having a less demanding protocol that would obtain the essential information even of it did not include the whole protocol.

4. An Illustration: A Homogeneous Approach to Standard Information vs The Most and Best You Can Get?

Below, I offer an example to illustrate the process; this is the approach developed at the National Opinion Research Center (NORC) for REACH-US (Racial and Ethnic Approaches to Community Health – United States) a multicommunity study of health-related attitudes and behavior in relatively deprived communities in the US. This is not intended to suggest that a particular approach, or sequence of approaches, is optimal, but merely to illustrate some of the possibilities. The areas involved are primarily urban; the diagram relates to these areas. [In two areas no address frame existed, and face-to-face interviewing was the first mode attempted in these communities.] The populations are defined by geographical area, based on census geographies (blocks, block groups, or tracts).

Multi-mode Address Based Sampling

Underlying this approach is the belief that coverage is a fundamental requirement of the survey. Thus the starting point is the address frame that covers (essentially) all the dwellings in the community. Evaluation work on the frame (the Delivery Sequence File) suggests that the coverage rate is above 97% in urban areas.

Of the three modes being used in this study (mail, telephone, and face-to-face interviewing) telephone is the cheapest. Consequently the first stage of data collection was based on telephone numbers that matched the selected addresses. This operation was carried out by CATI from a centralized telephone facility.

Where no response was obtained by CATI, or where no telephone number was found that matched the selected address, a mail questionnaire (self-administered questionnaire (SAQ)) was sent to the address. As selection of respondents within household is highly problematic, responses were requested from all qualified adults in the household. Respondents were given two other alternatives: calling a toll-free number on which an interviewer would be available to carry out the interview by phone, or return a very short screening questionnaire, after which an interviewer would call. These two alternatives returned the case to the telephone mode. Various incentives were built into this process.

Among those selected cases that failed to respond either by phone or to the mail questionnaire, a subsample of respondents was selected for field follow-up. In this final stage, the interviews were completed face-to-face. [This field stage was possible partly because the small geographical area of the communities made travel costs relatively low.]

5. Conclusion

Underlying this approach is a re-thinking of the survey operation. Hitherto there has been among methodologists a very strong concern about mode effects, built in large part on a not very large number of experiments comparing the responses of matched samples of individuals. These experiments neglect the broader question of the combination of coverage, nonresponse and data quality that are implied by the choice of a particular mode.

Our objective should be to obtain the best possible data for the population (taking into account not just sampling error, but also coverage error, nonresponse error, and data quality). If faced with the choice between getting no data for a selected individual (because of noncoverage, say) and getting data in a different mode, most users (and stakeholders) would almost certainly prefer the latter (as they should, unless the alternative mode were so egregiously deficient that adding in the data decreased the overall quality of the information). The dynamic iterative nature of the REACH M-ABS design exploits the strengths of all the approaches simultaneously within a single design framework.

Figure 5: An Illustration of a Multimode Address-Based Sampling Design