

SURVEY METHODOLOGY RESEARCH FOR IMPROVED HEALTH STATISTICS

James M. Lepkowski, Mick P. Couper, Nancy Mathiowetz, Roger Tourangeau, and Trivelloro Raghunathan, University of Michigan

James M. Lepkowski, Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48109

Key words: Disability, proxy reports, sensitive behaviors, audio-CASI, hierarchical models, Bayesian methods.

The University of Michigan Center for Excellence in Health Statistics seeks to improve the quality of survey data available to federal, state, and local policy makers for health promotion and disease prevention. Survey methods have been called as one of the great inventions of the 20th Century, and are widely used in health policy formulation for information about the health and health care utilization of the general population. Sound, accurate, and reliable data are required for policy analysts not only to assess the current state of health and health care utilization in the general population but also to develop models to predict trends and developments in the population's health. Improving the quality of health survey statistics leads to more effective health promotion and disease prevention policy. The Michigan Center for Excellence in Health Statistics (MiCEHS) is designed to conduct research on the survey process to improve the accuracy of data obtained from sample surveys.

The Center builds on research at the University of Michigan and its partner in the proposed Center, the Gallup Organization, which uses quantitative and qualitative methods to study the survey research process. Survey research methodology is at present a nascent discipline, investigating the survey process through interdisciplinary studies blending theory from social, behavioral, health, economic, statistical, and computer sciences into comprehensive theories of information collection and utilization. The MiCEHS will develop theory to explain behavioral and statistical features of the survey process, and generate empirical findings to test and refine the theory.

The MiCEHS has four principal aims:

1. Elaborate the existing infrastructure to support interdisciplinary survey methodology research in a setting where methodological practice and substantive application are joined to improve the quality of survey data and develop new statistical methods for health promotion and disease prevention.
2. Conduct interdisciplinary methodological research on the survey research process using,

when appropriate and convenient, ongoing survey data collection activities at the University of Michigan and the Gallup Organization as vehicles for research and laboratory studies.

3. Develop new statistical methods for analyzing survey data in collaboration with other researchers in the School of Public Health, the Gallup Organization, and the National Center for Health Statistics.
4. Promote educational outreach to health statistics researchers at the National Center for Health Statistics and elsewhere to inform about the latest research findings and to train in research methodologies in the survey research process.

The following describes the infrastructure and ongoing research activities at the MiCEHS.

Infrastructure

The MiCEHS includes several initiatives to promote the research activities by Center members. Two of these, an interdisciplinary seminar in health statistics and an NCHS data center, are discussed here.

An ongoing bi-weekly seminar provides an intellectual meeting ground of Michigan researchers and visitors who are making contributions in fields related to health statistics. The seminar meets throughout the academic year (September through May) and is organized around themes on important issues facing the field of health statistics. Themes can be methodological, soliciting participation of substantive researchers who are interested in methodological developments as contributors or discussants, or substantively motivated by measurement problems in health statistics. A video link between Ann Arbor and other locations is used to broaden the participation to include National Center for Health Statistics staff.

A second component of the MiCEHS infrastructure is an NCHS data center. NCHS has developed an administrative mechanism to increase access to confidential data resources through a data center in its Hyattsville offices. The MiCEHS is seeking NCHS approval to establish an offsite data center as part of the

MiCEHS in order to increase the range of topics that the proposed MiCEHS can address. For example, one of the three projects developed later will require access to data that are not part of the public use data released by NCHS at present.

This infrastructure is designed to supplement the primary research activities of the Center. The MiCEHS has begun with three research projects chosen by four program directors and the Center director. But a mechanism of renewal and innovation is needed to sustain a long-lasting contribution to health statistics. A mechanism has been established within the Center to review, approve, and fund future research projects. A system of review and approval jointly between the MiCEHS staff, NCHS staff, and a Center advisory committee will assure that the best research ideas are funded. There will be preliminary development to assess the feasibility of alternative research projects among MiCEHS senior staff, proposal review, and scoring by advisory committee and NCHS staff, and evaluation and selection of projects for funding.

Research Projects

Three research projects are underway in the initial phase of the MiCEHS. These illustrate the interdisciplinary character of collaborations among the Center's research staff. The projects address all three of the challenges that confront health research—the conceptual and measurement issues raised by emerging issues, the issues raised by new data collection technologies, and the questions raised by the application of new estimation techniques.

Project 1: Measurement of Persons with Disabilities

The primary focus of this investigation is the measurement of persons with disabilities. The project seeks to assess the effect of alternative questionnaire wording on reliability (simple response variance) and the effects of self and proxy reporting on the reliability of responses.

The six disability questions included in the long form of the 2000 decennial census are the basis for the experimental design. Alternative versions of the six questions have been designed to investigate the impact of alterations on response reliability. Alterations include:

- the use of dichotomous response v. a 5 point response scale
- separating compound questions into several shorter, simpler questions

- separation of the means by which the task is performed from the difficulty associated with performance of the task.

The study will consist of two successive interviews in 800 households. In order to increase the incidence of identifying persons with disabilities, only those households in which at least two people ages 40 and older reside will be eligible for the interview. Each household will be randomized to a question treatment for a sensory impairment question and separately randomized to a version of an activity limitations questions. In the first interview, a randomly chosen respondent will be asked to respond to the questions concerning him- or herself and then to report about one additional person in the household age 40 and older ('other'). Approximately two weeks later, the household will be recontacted. In half of the households we will interview the same person who responded in the initial interview; in the remaining cases we will interview the 'other' person for whom data was collected in the first interview.

One major concern with respect to designing questions concerning the measurement of persons with disabilities is that we have little understanding of respondents' perception of what constitutes a disability and how that does or does not correspond to theoretical models used by researchers. In an attempt to understand respondents' perceptions of disabilities, each respondent will be read a vignette of a person and asked as to whether he or she perceives the person described in the vignette to have a disability. The vignettes vary on a number of characteristics including the nature of the disability (e.g. physical, emotional, addiction), the persons' gender, whether or not the person is currently employed, the use of devices or medication, and other individual awareness of the persons' condition or impairment.

The design permits the assessment of simple response variance across various self-proxy treatments (self-self response; proxy-proxy response; and self-proxy response). In both waves of the study, respondents will be randomized to a vignette (different vignettes in each wave) and asked to respond to the perception questions with respect to the vignette description.

Tables 1 and 2 present the factorial design for the experiment. Due to the nature of the question concerning sensory impairments, we have treated that question differently than the remaining five questions which concern functional limitations and participation. Table 1 presents the design related to the single

question in the decennial Census concerning vision or hearing impairments; Table 2 presents design features for questions on physical limitations, learning, remembering or concentrating limitations, activities of daily living, going outside the home alone, or working at a job or business.

Table 1. Experimental versions for vision and hearing (sensory impairment) question

Factor	Version			
	A	B	C	D
Response scale	Y/N	5 pt.	Y/N	5 pt.
Shorter, simpler q's	No	Yes	No	Yes
Description	Std	Concrete	Std	Concrete

Table 2. Experimental versions for activity limitation questions

Factor	Version				
	E	F	G	H	I
Response scale	Y/N	5 pt.	Y/N	5 pt.	5 pt.
Shorter, simpler q	No	No	Yes	Yes	Yes
Task performance	None	None	None	None	Yes

The design of the Wave I and Wave II questionnaires was completed in early July. The Gallup Organization is collecting data via telephone and computer assisted interviewing. A small pretest will be followed by an interviewer debriefing, interviewer training for the main study, and data collection starting in August, 2000. We anticipate completion of both waves of data collection by early November, 2000.

Project 2: Obtaining Accurate Self-Reports of Sensitive Behaviors

A variety of health conditions carry personal or social stigma and, as a result, may be subject to social desirability bias in survey data collection. Social desirability biases are apparent in the over-reporting of positive behaviors (for example, voting in elections) and the underreporting of negative ones (such as illicit drug use). Health surveys include many examples of both types: socially undesirable behaviors related to health (e.g., drug and alcohol use, smoking, risky sexual practices, abortion), and socially desirable behaviors (e.g., healthy eating, exercise, dental visits).

One particular data collection technique that may improve the quality of information obtained on

sensitive topics audio computer assisted self-interviewing (audio-CASI). In audio-CASI, a respondent hears the questions over headphones attached to a laptop computer from a digitized recording of the question reading and enters responses directly into the computer.

The technique shows substantial promise in reducing social desirability effects for behaviors of interest to health survey researchers (Johnston and Walton, 1995; O'Reilly et al., 1994). Turner et al. (1998a; see also 1998b) embedded an experimental comparison of a paper and pencil self-administered questionnaire (SAQ) with audio-CASI for the most sensitive items on the National Survey of Adolescent Males (NSAM), and reported an almost fourfold increase in the number of respondents reporting male-male sexual contact, from 1.5 percent for paper and pencil to 5.5 percent for audio-CASI. Tourangeau and Smith (1996; see also 1998) compared computer assisted personal interviewing (CAPI), text-CASI, and audio-CASI, and found that across a wide variety of questions involving sexual behavior and drug use, audio-CASI and "text-CASI" (in which respondents recorded answers on paper rather directly into a computer) generally yielded higher levels of reporting than interviewer-administered computer assisted personal interviews.

While these initial studies have led to the widespread adoption of audio-CASI in large-scale national surveys on a variety of sensitive topics (e.g., the National Household Survey on Drug Abuse, National Survey of Family Growth, and National Survey of Adolescent Males), only one study (Tourangeau and Smith, 1996, 1998) has directly compared audio-CASI to text-CASI. Audio-CASI has additional cost and effort relative to text-CASI for large-scale health surveys. It is thus important to explore the relative advantages of different approaches to self-administration.

Survey-based studies of audio-CASI have focused on the increased privacy offered by audio-CASI with respect to other people present in the interview setting. They have largely ignored the potentially biasing effect of the voice used in the audio-CASI device. Audio-CASI may produce *more*, not less, social desirability bias than text-CASI, because additional social cues are provided to the respondent in the form of the interviewer's voice.

Ongoing research is addressing human-computer interaction (HCI) issues with computer-assisted interviewing (CAI), but to date little or no research has focused on the design and implementation of CASI

instruments. Given the increasing use of methods such as audio-CASI in a wide variety of health survey and clinical settings, Project 2 addresses two key hypotheses that might explain the diverse reactions to CASI use in surveys: a) the dominant mode of delivery of the question (text or audio), and b) the presence of the interviewer.

One mechanism by which audio-CASI is hypothesized to improve the quality of self-report is the increased privacy it affords respondents with respect to other people who may be present during the interview. A second mechanism may be the social presence of the live or virtual interviewer.

The effects of audio-CASI relative to text-CASI are hypothesized to be situation-dependent. In a situation where other people *cannot* disturb the respondent's privacy, text-CASI is expected to obtain reports that are at least as "good" (*i.e.*, lower reports of socially desirable behaviors and higher reports of undesirable behaviors) as audio-CASI. However, in a situation where other people do potentially intrude on the respondent's privacy, it is expected that the effect of the virtual voice on audio-CASI will be negated by the enhanced privacy offered by this method. Audio-CASI should therefore yield reports that are more accurate than text-CASI.

To test the relative effectiveness of audio-CASI and text-CASI in reducing social desirability effects, the project is implementing in a laboratory setting a 2 x 2 design crossing the degree of privacy (high versus low) and the mode of administration (text-CASI versus audio-CASI). All conditions will be run in a laboratory setting to permit more carefully controlled measures of the hypothesized independent variables. While this feature of the design may raise questions of external validity, it allows full control of the experiment treatment. If the expected effects are found, later studies will attempt to replicate the findings in a field setting.

In the high privacy conditions, the interviewer will introduce the respondent to the audio- or text-CASI instrument and then leave the room. The respondent will be left alone in a private room for the entire time he or she is completing the survey instrument. The interviewer will be available outside the room should the respondent need assistance. In the low privacy conditions, both the interviewer and a laboratory assistant will be present as the respondent is completing the interview. The interviewer will remain in the room with the respondent during completion of the CASI questionnaire; in addition, a research assistant will enter and leave the room several times during the interview

on the pretext of checking that the equipment is working correctly.

A variety of sensitive health-related items will be included in the instrument. These will include questions about both socially undesirable behaviors (*e.g.*, number of sex partners, unprotected sex, drug and alcohol use, smoking, etc.) and socially desirable behaviors (*e.g.*, exercise, dental visits, diet). To maximize power for items that may be rare in the general population, subjects will be recruited in high-risk categories on several of these variables.

Following completion of the experimental questionnaire, subjects will answer debriefing questions designed to assess (a) how sensitive they considered the different questions to be; (b) how concerned they were about other people finding out how they answered the questions; (c) how much privacy they felt the method of interviewing afforded them; (d) how comfortable they were with the interviewing method; and (e) what alternative method of interviewing they would have preferred. In addition, a series of questions will be designed to elicit their awareness of the social presence of the virtual interviewer by asking about voice and other inferred characteristics of the interviewer. These measures will be used to supplement the actual behavioral reports, and will serve in part as a check of the effectiveness of the experimental manipulation.

A total of 200 volunteer subjects will be recruited for the experiment, yielding 50 subjects per cell.

Project 3: Combining Contextual and Individual Data from Multiple Sources

Scientific understanding of the relationship among health factors that lie at the individual level and the level of social and spatial aggregates has been severely hampered by the lack of analytic tools. With the widespread availability of those tools, though, it is clear that an important source of data about health has not produced the kind of contextual information needed to understand the interplay of individual and groups.

One source of contextual information that has not been fully exploited is aggregation of survey data themselves to higher level sampling units from which the survey subjects were selected. Many health surveys on different health topics are conducted in the same primary sampling units, yet there is little opportunity to link the sampling unit across sample designs. For example, the National Health and Nutrition Examination Survey (NHANES) is conducted in a sample of the National Health Interview Survey's (NHIS) primary sampling units. NHIS data could be used to estimate neighborhood structural characteristics,

such as ease of access to a health care facility at the PSU level, which then can be related to individual health risk factors such as cholesterol level. What is lacking are methods to exploit the existence of coordinated survey design for the measurement of ecological influences.

The objective of this project is to use multi-level or hierarchical models to develop statistical methods and associated software to bring community and neighborhood foundations of health and development into analysis of individual health characteristics.

Recent developments in Bayesian computation (e.g., Markov chain Monte Carlo methods) have made it possible to apply hierarchical models to both continuous and categorical outcome data (Geman and Geman, 1984; Gelfand and Smith, 1991). Further, Bayesian methods have the desirable property that they can use more of the information available more efficiently than traditional frequentist procedures. The proposed project will use the Bayesian computational framework to combine information from neighborhood and individual characteristics in a sample survey to examine individual health outcomes through a set of random effect hierarchical models. The random effects estimated from one level of the model are used as predictors in the next level of the model.

Let x_{ij} denote an indicator variable taking the value 1 if the person i in neighborhood j is below the federally defined poverty level, based on a detailed assessment in a large survey such as NHIS. A random effect logistic regression model may be used to specify the relationship between x_{ij} and a set of predictors such as region or, urban residence, denoted as v_{ij} , as follows:

$$x_{ij} \sim \text{Bernoulli}(1, \theta_{ij})$$

$$\log \text{it}(\theta_{ij}) = \beta_j + (v_{ij} - \bar{v}_j)\lambda$$

Here the β_j are random regression coefficients, the adjusted community-level log odds of being below the poverty level, and are assumed to be normally distributed with mean 0 and variance τ^2 . λ is a vector of fixed effect regression coefficients. Suppose that y_{kj} is an individual health outcome of interest such as blood glucose for subject k from the same neighborhood, but y_{kj} is measured in another survey using the same neighborhoods, or primary sampling units. A second-stage model regresses this health

outcome on the unobserved random effect (β_j) and individual-level variables u_{kj} ,

$$y_{kj} = \alpha_o + \alpha_1 \beta_j + u_{kj} \gamma + \varepsilon_{kj}$$

Here ε_{kj} are assumed to be normally distributed with mean 0 and variance σ^2 . The object of the inference is α_1 , the adjusted effect of the neighborhood characteristic.

This second-stage model is suitable for normally distributed health outcomes. A logistic regression model could be used for binary outcome variables such as presence or absence of a condition; a polytomous or multinomial regression model may be used for nominally or ordinal-scaled outcome variables such as self-rated health or type of health insurance coverage. The method could readily be adapted to handle count variables such as the number of visits to a doctor's office through a Poisson model.

Gibbs sampling and other Markov chain Monte Carlo algorithms will be used to construct posterior distribution of the parameter of interest, α_1 and the other parameters in the model. As a first step, algorithms will be developed for drawing values from first stage of the model, conditional on the parameters in the second stage of the model and on the data. Next, procedures will be developed for drawing values from the posterior distribution of the parameters of the second stage of the model, conditional on the first stage parameters and on the data. These two sampling algorithms will be combined into a single general-purpose software system to implement the procedure. The software procedure will be implemented in SAS using facilities such as the macro language, PROC IML (and interactive matrix language), and the SAS ASSIST features to present screens that allow users unfamiliar with the complexities of Bayesian methods, Gibbs sampling, and Markov chain Monte Carlo methods to specify substantively suitable models.

The project will also explore whether "neighborhood" characteristics estimated using NHIS data may become part of the public use NHANES (or NSFG) files with suitable random recodes of the primary sampling unit characteristics to assure confidentiality. Thus, analysts outside of NCHS will have access to the random effects coefficients representing neighborhood or primary sampling unit characteristics that would ordinarily be inaccessible.

References

- Gelfand, A. E., and Smith, A. M. F., (1990). "Sampling Based Approaches to Calculating Marginal Densities," *Journal of American Statistical Association*, 85, 398-409.
- Geman, S., and Geman, D., (1984). "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Johnston, J., and Walton, C. (1995). "Reducing Response Effects for Sensitive Questions: A Computer-assisted Self Interview with Audio," *Social Science Computer Review*, 13, 3, pp. 304-319.
- O'Reilly, J.M., Hubbard, M., Lessler, J., Biemer, P.P., and Turner, C.F. (1994). "Audio Computer Assisted Self-Interviewing: New Technology for Data Collection on Sensitive Issues and Special Populations," *Journal of Official Statistics*, 10, pp. 197-214.
- Tourangeau, R., and Smith, T. (1996). "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context," *Public Opinion Quarterly*, 60, pp. 275-304.
- Tourangeau, R., and Smith, T. (1998), "Collecting Sensitive Information with Different Modes of Data Collection," in Couper, M.P., *et al.* (eds.) (1998), *Computer Assisted Survey Information Collection*. New York: Wiley.
- Tourangeau, R., Smith, T.W., and Rasinski, K. (1997). "Motivation To Report Sensitive Behaviors In Surveys: Evidence From A Bogus Pipeline Experiment," *Journal of Applied Social Psychology*, 27, 209-222.
- Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., and Sonenstein, F.L. (1998a). "Adolescent Sexual Behavior, Drug Use and Violence: Increased Reporting with Computer Survey Technology," *Science*, 280 (May 8): 867-873.
- Turner, C.F., Forsyth, B.H., O'Reilly, J.M., Cooley, P.C., Smith, T.K., Rogers, S.M., and Miller, H.G. (1998b). "Automated Self-Interviewing and the Survey Measurement of Sensitive Behaviors," in Couper, M.P., *et al.* (eds.) (1998), *Computer Assisted Survey Information Collection*. New York: Wiley.