WORKSHOP ON MICROSIMULATION MODELS FOR SURVEYS

Abstracts

Overview of Simulation Models and a Simulation Model for NHIS Field Operations and Cost Estimates
Bor-Chung Chen, U.S. Department of Transportation

Discrete-event simulation modeling has become the most commonly used tool for performance evaluation of stochastic dynamic systems in science and engineering. The field operations of surveys can be classified as one of these stochastic dynamic systems. Simulation modeling provides, if feasible, flexibility to build either aggregate or detailed models. This presentation gives an overview of simulation modeling methodology and describes the simulation and modeling of simplified field operations for NHIS (National Health Interview Survey). We will describe simulation study steps needed to develop a simulation model that is a valid representation of an existing/proposed real system. We use the 2004 NHIS CHI (Contact History Instrument) data for the input modeling of the simulation. From this study, we have shown that simulation modeling can be used for optimizing the field operations by setting the controllable parameters before a decision is made and implemented. The cost savings might be enormous and would not be at the expense of the response rate.

Simulation Models to Inform Health Policy
Carolyn Rutter, Group Health Research Institute

Simulation models are increasingly used to evaluate and inform healthcare decisions. Researchers who develop these models aim to integrate and synthesize short-term outcomes and results from multiple sources to predict the long-term clinical outcomes and costs of different healthcare strategies. Policy makers, in turn, can use the predictions generated by disease models together with other evidence to make decisions related to healthcare practices and resource utilization. This presentation will provide an overview of disease simulation models, including both cohort and microsimulation models, touching on model building, selection of model parameters to reproduce observed or expected results (calibration), model validation, and issues related to reporting and interpreting findings (sensitivity analyses, reporting of variability, and model transparency). These ideas will be demonstrated based on our model for colorectal cancer. This work was supported by NCI U01 CA52959.

Some Cost-Modeling Topics for Prospective Redesign of the U.S. Consumer Expenditure Surveys
John L. Eltinge and Jeffrey M. Gonzalez, U.S. Bureau of Labor Statistics

The U.S. Consumer Expenditure (CE) Surveys are designed to obtain detailed information on expenditures by the U.S. noninstitutionalized civilian population. At present, there are two survey components: a diary, intended to capture expenditure data for small and frequently purchased items; and an interview, intended to capture information on other expenditures, as well as income and assets. The diary covers two consecutive weeks, while the interview involves five waves of data collection, with approximately three months between waves.

BLS management is currently considering a wide range of options for redesign of the CE in the presence of several changing factors in the economic and data-collection environments. For each option, evaluation will depend heavily on empirical information on fixed and variable cost components, as well as several distinct dimensions of data quality and operational risk. The applicable empirical information is often difficult to collect, and may be subject to confounding or aggregation effects. In addition, observational data on cost and quality may be affected by nonrandom assignment of certain tasks in fieldwork. Consequently, it may be useful to supplement directly observed empirical information with data obtained through microsimulation methods.

Following presentation of some general background, this talk will highlight two areas in which microsimulation work may be of special interest. The first area involves the prospective integration of survey data with administrative-record data, e.g., expenditure information that might be available (with appropriate informed consent) from retailers or financial intermediaries. For this work, an effective balance of cost, quality and risk would require evaluation of costs related to data acquisition and integration, systems development and maintenance, and human resources; quality measures involving the rates and informativeness of incomplete-data patterns, definitional issues, and temporal and cross-sectional aggregation; and risk factors related to the prospective sources of administrative data.

The second area involves the use of responsive design methods, in which information collected earlier in the survey process (e.g., on a previous wave, or in the initial minutes of a specified wave) may be used to determine the specific detailed expenditure items that CE would collect from a given sample consumer unit. Issues of primary interest center on the components of cost attributable to unit initiation, wave initiation and incremental minutes of data collection; components of data quality related to burden, framing effects and recall effects; and components of risk arising from standard nonresponse and reporting-error processes, as well as possible noncompliance of field representatives with responsive-design directions.

Optimizing and Simulating CATI Call Scheduling

Mike Hidiroglou, Francois Laflamme, H. Choudhry and Yves Bélanger, Statistics Canada

One of the main increasing challenges for Statistics Canada is to collect cost-effective data while maintaining a high level of quality. Paradata research has been useful in improving the current data collection process and practices. The research carried out with paradata suggested that collection resources are currently not always optimally allocated with respect to the assigned workload and the corresponding expected productivity.

We present two approaches to optimize CATI data collection. In the first approach, a micro simulation system that mimics the collection process can be used to reduce data collection costs In the second approach, models are developed to predict the probability that a telephone call result in a completed questionnaire as a function of time of day, and resources spent to date. The estimated parameters are input into a loss function that optimizes call scheduling subject to constraints.

The presentation begins with a brief description of the cost-efficient framework followed with an overview of the simulation project. The second part of the presentation mainly focuses on models to predict survey productivity as a function of time of day, and resources spent to date and on the parameters estimated from these models used as input to optimize call scheduling. The link between the optimization and simulation projects will also be discussed.

Standard Errors and Uncertainty in Agent-Based Models
Georgiy Bobashev, Research Triangle Institute International

Agent-Based Models (ABMs) can be used to quantify future risks by projecting observable behavior into the future. This can be achieved by simulating a hypothetical longitudinal study based on cross-sectional data and estimating quantities on dynamic risks (e.g., relative hazard). Such an approach, however, requires assessment of the variation of the estimates, which would naturally have a higher variance than would be achieved in a real longitudinal study. We present a methodology that considers rigorous statistical measurements such as standard errors and uncertainty associated with the fact that the analyzed longitudinal data are a projection of the cross-sectional survey. We illustrate the use of our approach in simulated and real studies.

Developing a Microsimulation Model of a Federal Survey
Ben Klemens, US Census Bureau

The problem of developing a model of surveys is not necessarily about working from scratch, but tying together results from the many involved fields. Cognitive testers have gathered data on how respondents respond to survey questions, via telephone, Internet, or in person. The responsive design literature has gathered knowledge on how to map learned characteristics about the population to mobilization of resources on the ground. For a survey of our invention regarding a population of our invention, standard survey statistics report the bias and variability of the final survey results. I will describe the requirements of a framework that unifies the teachings from such disparate fields.