Title: Benefits and Challenges in Using Paradata

Date and Time: April 18, 2016
12:30pm to 3:30pm

Moderator: Mike Fleming

Sponsor: WSS Methodology Section

Location: Offices of Mathematica-MPR 1100 1st Street NE, 12th Floor, Washington DC 20002. Once in the building, inform the receptionist at the first floor lobby that you are visiting Mathematica for a WSS seminar. Then, take the elevators to the 12th floor and tell the Mathematica receptionist that you are attending the WSS seminar. Please call Mathematica's main office number (202 484-9220), if you have trouble finding the building.

By Metro: Take the Red Line to either the NoMa-Gallaudet U (used to be called New York Ave) Station or Union Station. From the NoMa-Gallaudet U Station, follow signs to exit at M Street. Then walk 1 block west on M street and 2 blocks south on 1st Street NE (the building will be on your right). From Union Station, walk north along 1st Street NE for about 4-5 blocks until you reach L Street (the building will be on your left after crossing L street).

By Car: Pay parking is available in the building parking garage, which is located 1 block east of North Capitol on L Street NE.

Guest List: To be placed on the attendance list (in-person or webex), please RSVP to Alyssa Maccarone at AMaccarone@mathematica-mpr.com or (202) 250-3570 at least 2 days in advance of the conference. Provide your name, affiliation, and contact information (e-mail is preferred). Once on the attendance list with webex preference, you will be provided with information about webinar.

Lunch option: Attendees may arrive early to have lunch nearby. Local lunch options may be found through: http://www.nomabid.org/wp-content/uploads/2011/02/FINAL_NeighborhoodGuide.pdf. You may also bring your own lunches to the seminar.

Following the seminar, **snacks and refreshments will be served**, encouraging the **attendees** to continue questions and discussions on the talks.

Schedule

Time	Speaker	Affiliation	Point of Contact
12:30	Mike Fleming		charles.fleming@bhox.com
12:40	Brady West	University of Michigan	bwest@umich.edu
1:05	Emilia Peytcheva	RTI International	epeytcheva@rti.org
1:30	Stephanie Coffey	US Census Bureau	Stephanie.Coffey@census.gov
1:55	Intermission		
2:10	Jason Markesich and Shawn Marsh	Mathematica Policy Research	JMarkesich@mathematica-mpr.com
2:35	James Wagner	University of Michigan	jameswag@umich.edu
3:00	Floor Discussion		

Title: Interviewer Observations in the National Survey of Family Growth: Lessons Learned and Unanswered Questions

Brady West

Abstract:

In this presentation, I will discuss several studies of the quality and utility of interviewer observations in the National Survey of Family Growth, ongoing research related to improving their quality, and future research directions for these types of observations. Existing research suggests that these observations can be useful for responsive design and nonresponse adjustment purposes, and that in some cases the observations are more effective than linked commercial data. However, there is a great deal of room for improvement in terms of the collection and analysis of these observations, and the presentation will touch on necessary future research in this area.

Title: Modelling Paradata for Interviewer Monitoring

Emilia Peytcheva

Abstract:

Interviewer's performance is essential to data quality. Paradata play a critical role in monitoring and improving interviewer performance, but are not as useful as they could be, in their raw form. For example, even in a centralized CATI survey, cases are not randomly assigned to interviewers; some interviewers only work on evenings, others may not be permitted to call previous refusals, etc. There are also multiple aspects of interviewer performance, ranging from success rate in gaining participation to administering the questions as intended.

We describe a system for interviewer performance monitoring that provides an automated and standardized means to monitor and improve interviewer performance in a call-center environment. Two types of multivariate models are of interest. First are models for predicted interviewer response rate, as a function of type of sample called in a shift, language spoken by the sample member, number and types of cases attempted, etc. These models were started on a study in early 2010 and are now used on multiple surveys. Versions of these models have also been implemented to predict refusal conversion rate, as a function of call history paradata.

In addition, we are currently developing models to incorporate coded interviewer behavior from monitored recorded interviews and interactions. These models aim to provide interviewer-level measures of interviewer adherence to the survey protocol and to allow prompt feedback to interviewers before such behavior can impact the measurement properties of the data.

Title: Using System Paradata to Target and Evaluate Data Collection Operations

Stephanie Coffey

Abstract:

Paradata is being used increasingly for both understanding data collection operations as well as making changes to data collection. Paradata can be either interviewer-generated or system-generated, depending on the data collection operation and the system being used. This talk will discuss three examples of the use of paradata in the National Survey of College Graduates (NSCG). The NSCG is a 6-month long sequential multimode survey, cycling through web self-response, paper self-response, and finally, computer-assisted telephone interviewing for nonresponse followup. Therefore, the vast majority of our paradata is system-generated paradata.

In this talk, I will present three illustrations of how paradata is being used in the NSCG. First, I will discuss an ad-hoc intervention implemented during the 2013 NSCG to respond to an unexpected performance issue in the web instrument. Next, I will present a simple illustration of how we are evaluating CATI as a driver of paper or web response. Finally, I will demonstrate how we are using intelligent mail barcoding (IMB) data to allocate resources to cases more efficiently.

Title: Best Practices for Implementing a Paradata Warehouse

Jason Markesich and Shawn Marsh

Abstract:

Paradata are a powerful tool for increasing the effectiveness of data collection. To ensure high quality survey data and low-cost data collection, Mathematica developed the Paradata Warehouse, a centralized, standardized repository that will contain paradata for all of our projects. New project data are added to the warehouse, and are accessible, on a daily basis. Survey managers can conduct cross-project and/or cross-instrument analysis of paradata to inform design and budgeting decisions. Because the warehouse tracks hundreds of variables across multiple projects, we have incorporated business intelligence tools and data visualization software to help users understand the story their data is telling. As part of our adaptive design initiative, our survey managers can analyze these paradata and tailor data collection strategies to respond quickly to conditions “in the field” to maximize efficiency and improve quality.

This presentation discusses some best practices for implementing and deploying a paradata warehouse that collects and organizes information in a manner that is easily accessible and meaningful to managers. We begin by discussing the steps we have taken to ensure the accuracy of the data in the warehouse through the standardization of status and charge codes, metadata management, and testing and reconciliation processes. Next, we present some data visualization techniques that help engage users by providing them with answers quickly, and in an easy-to-understand format. We conclude by offering some tips on how to develop a strategy that will help facilitate user adoption of a paradata warehouse.

Title: Estimating Response Propensity Models During Data Collection: Challenges and New Approaches

James Wagner

Abstract:

Response propensity models have been used to create inputs to adaptive survey designs. These inputs may be needed during data collection as triggers for design decisions. However, predictions from response propensity models can be biased when fit on a daily basis during data collection using the incoming data. One solution to this problem might be to fix the estimated coefficients from these models using estimates from previous data collections. This approach has two disadvantages. First, a suitably matched data collection must exist. Second, it is not sensitive to temporal changes that may change estimated coefficients. Bayesian logistic regression models, with informative priors, may address these weaknesses. This presentation examines some of the challenges of working with incoming streams of paradata and suggest potential solutions.