

List of titles and abstracts for one-day conference on “Paradata: From Survey Research to Practice”
10:30 - 17:30, 26 June 2014, Royal Statistical Society, London
www.southampton.ac.uk/paradata_event

List of Titles and Abstracts
(in alphabetical order of surname of speakers)

Using paradata to improve the quality of web surveys. Some examples and applications

Mario Callegaro

Google UK Ltd.

Paradata in web survey have been used to improve the quality of data collection, from studying nonresponse, to improve question wordings, to find problems and issues during the completion of a questionnaire. Although paradata in web surveys are “relatively” easy and cheap to collect, the process and analysis effort is extremely complex and time consuming if not guided by precise research questions. In this presentation a taxonomy of three paradata types for web surveys is proposed: contact management, device type, and questionnaire navigation paradata. For each type some examples are presented in order to provide the audience with concrete applications of using paradata to improve data quality in web surveys. The goal is to provide a state of the art of what paradata are possible to collect and how to answer specific research questions in terms of data quality on web surveys.

Modelling final outcome and length of call to improve efficiency in call scheduling

Gabriele Durrant, Olga Maslovskaya, and Peter W.F. Smith

Southampton Statistical Sciences Research Institute, University of Southampton

For interviewer administered surveys many survey agencies nowadays routinely collect call record data. Examples of such data may be recordings of the day, time and outcome of each call or visit and, for face-to-face surveys, information collected by the interviewer such as physical and social characteristics of the selected housing unit and neighbourhood. Researchers have increasingly become interested in how best to use and analyse such information. It is hoped that a better understanding of the calling patterns and the mechanisms leading to particular call sequences will help to improve data collection.

More specifically, for statistical agencies, investigating time and effort into repeated calls and follow-ups is very resource-intensive. From a survey management perspective, it seems desirable to avoid long unsuccessful call sequences to improve efficiency. The aim then is to identify early on in the data collection process cases prone to long and unsuccessful call sequences. This paper models call record data predicting final call outcome and length of a call sequence early on in the data collection process. Separate logistic and joint multinomial models for the two outcomes are considered and the

models account for the clustering of sample cases within interviewers. Of particular interest is to identify good explanatory variables that predict final outcome and length of a call sequence, in particular characterising long unsuccessful call sequences. Findings so far indicate that outcomes of the initial call attempts are highly predictive but also other variables such as interviewer observations. Further research questions that we aim to address in this study are: how can predictors best be incorporated into the model (e.g. as summary statistics or as individual outcomes)?; how predictive are the models?; does their ability to predict the variable of interest improve if more and more call record data are available (e.g. for later calls; or for later waves in a longitudinal study)?; how can these models best be used in adaptive and responsive survey designs?

The study uses data from a large-scale longitudinal survey in the UK, Understanding Society. Implications for survey practice and fieldwork procedures are discussed.

Use of paradata derived from vocal properties of interviewers and respondents in researching survey participation

*Frauke Kreuter**, *Fred Conrad*, *Jose Benki*, *Alexandra Birg*, *Susanne Hüttinger*, and *Helmut Küchenhoff*
*University of Michigan, University of Munich, and Institute of Employment Research (IAB)

In their decisions to participate in a telephone survey, potential respondents likely take into consideration how the conversation with the interviewer might look like. During the recruitment phase respondents experience a sample of such conversation and gather information on the interviewers speech pattern, sound, and their interaction with the respondent. Using a total of 1380 survey invitations from five different surveys conducted at the Survey Research Center at the University of Michigan, Conrad et al. show that interviewer speech rate and pauses do indeed affect participation. This paper expands this work, by examining if interviewers' adaptation to the respondents vocal characteristics does increase willingness to participate. We found a positive association between convergences in rate of speech and participation, though not for convergence in pause duration. Increasing synchrony in vocal pitch also increased the chance of agreement.

Collecting and publishing paradata on the German Internet Panel

Ulrich Krieger and *Annelies Blom*
University of Mannheim

The increase in the collection of paradata in recent years is based on the assumption that they have predictive power to explain survey errors in coverage, nonresponse, measurement or coding. In addition, their use is frequently justified by the easy with which they can be collected during the survey process.

In probability-based online panels, the potential for collecting paradata is large – spanning paradata on the recruitment process in general and, at each wave, on the respondents' computer settings, nonresponse patterns, question time stamps and key strokes. While the collection of paradata is typically connected with low additional costs, the publication of some types of paradata requires detailed data preparation. Furthermore, warranting respondents' anonymity and informed consent to data collection demands from researchers that they deal carefully with these data; this might inhibit the publication of some potentially available paradata.

This presentation showcases how the German Internet Panel (GIP) faces the challenge of paradata in terms of their collection and publication. We demonstrate how some types of paradata are easily made available to the research community, while others are associated with high additional costs and ethical concerns.

The GIP (http://reforms.uni-mannheim.de/internet_panel/home/) is based on a random probability sample that is recruited offline and represents both the online and offline population aged 16 to 75 in Germany. It was first set up in 2012 with a refresher sample in 2014. Once online, panel members are interviewed every two months via web-questionnaires on social, political and economic topics. The GIP data are published via the GESIS Data Archive for the Social Sciences within 6 months of their collection.

Sequence analysis as a graphical tool for investigating call record data

Olga Maslovskaya, Gabriele Durrant, and Peter W.F. Smith

Southampton Statistical Sciences Research Institute, University of Southampton

In recent years, analysis of call record data has found increasing attention. Survey agencies hope a better understanding of such data may inform improved interviewer calling practices, identify more difficult cases and unusual interviewer behaviour earlier on in the data collection procedure and may provide strategies for improved nonresponse adjustment methods. Although survey researchers have become increasingly interested in understanding and improving the process of data collection, to date, analysis of interviewer calling behaviour is still limited. It is often not clear, how best to analyse such data, in particular since call data can be large and may exhibit complex hierarchical and time-dependent data structures.

This paper introduces sequence analysis as a simple tool for investigating call record data to better understand and improve survey processes and designs with practical guidelines for survey managers. We will also cover the use of sequence analysis for longitudinal surveys. Sequence analysis offers a nice way of visualising, displaying and summarising the normally quite complex call record data. Sequence analysis tools have only more recently been introduced to survey methodology. A few papers exist that investigate the use of sequence analysis for improving nonresponse adjustment methods based on call record data. They conclude, however, that sequence analysis on call record data do not lead to significant improvements of nonresponse adjustment methods. Here, the method is used to inform survey management for adaptive and responsive survey designs. Sequence analysis is combined with clustering, optimal matching and multidimensional scaling. The sequence analysis method is applied to call record data from the UK Understanding Society survey. Implications of the findings for survey practice are discussed.

Using paradata to identify response styles and respondent profiles for adaptive survey design

Barry Schouten

Statistics Netherlands (CBS) and Utrecht University

Adaptive survey designs originate from a focus on nonresponse error and from trade-offs between proxy measures for nonresponse bias and survey costs. The rationale behind the designs is that survey target populations are diffuse in their response behaviour to different survey design features. The most prominent design feature in survey design decisions is the survey mode as it varies strongly in both costs and accuracy. The survey mode is, however, known to produce method effects that are broader than just differences in nonresponse error, and a single-minded focus on nonresponse in

adaptive survey design is, therefore, too naïve and simple. To become applicable in mixed-mode surveys, the adaptive survey design framework needs to be extended to measurement error.

There have been some recent attempts in the literature to develop a framework in which adaptive survey designs account for both mode-specific nonresponse error and measurement error. For a survey with only one key variable or with strongly related key variables, such a framework is easily constructed, as one may focus directly on minimizing method effects on a single survey variable. One would still need to have estimates of the method effects for different subpopulations, but the optimization problem is one-dimensional. For a survey with multiple and diverse key variables, such a framework is not easily constructed, because the optimization problem is multi-dimensional. In such settings, it has been suggested to adapt the design to forms of mode-specific answering behaviour that lead to measurement error on many survey variables simultaneously, i.e. to response styles. Response styles are deficiencies in the answering process that persist throughout a significant part of the questionnaire. In the literature there is a vast amount of papers that investigate one or more response styles. Most of them relate the response styles to the cognitive steps in the answering process and many relate them to the survey mode.

Before an adaptive survey design framework can be developed that employs mode-specific response styles, it has to be established to what extent response styles can be predicted from characteristics of the survey questionnaire and the respondent. In the paper, I discuss the design of an experimental study linked to the Dutch Labour Force Survey and to the Dutch LISS-panel. With the experiment, the utility of paradata (time measurements, audit trails, interviewer observations) is investigated in identifying mode-specific response styles and the deduction of respondent and questionnaire profiles. The results from the experiment form the input to research on adaptive survey design.

Using R-indicators for adaptive follow-up in longitudinal studies

Natalie Shlomo and Ian Plewis

Social Statistics, University of Manchester

Representativity or R-indicators measure the degree to which respondents and non-respondents differ from each other (the contrast) and go beyond response rates alone. R-indicators are based on the variation of response probabilities estimated through a response propensity model conditional on auxiliary information and paradata available to both respondents and non-respondents. R-indicators can be further decomposed into partial R-indicators which determine relevant population sub-groups that are contributing the most to the lack of representativity and can be targeted in data collection. For the case of longitudinal studies, there is a wealth of information and paradata from previous waves for developing response propensity models and R-indicators to implement targeted follow-up strategies in subsequent waves. The longitudinal survey design is generally more complex and hence response propensity models and confidence intervals of R-indicators need to be adapted to deal with clustered and weighted survey data. An application for adaptive follow-up in the longitudinal UK Millennium Cohort Study is presented.

Using response latency paradata to evaluate indicators of data quality

Patrick Sturgis, Gosia Turner, and David Martin

University of Southampton

According to Krosnick's influential account, survey satisficing occurs when a respondent decides to use a lower level of cognitive effort in order to provide a satisfactory but less accurate answer than

would have been produced if a greater amount of effort had been expended on the task. Satisficing theory has rapidly become the dominant framework in survey methodology for understanding response quality, with an increasing number of studies seeking to understand the causes and consequences of the decision to satisfice. However, the utility of commonly used empirical indicators of satisficing for assessing the accuracy and completeness of response data is open to question, because the prevalence of these indicators is related to a range of factors, in addition to a respondent's decision to satisfice. In this paper we use response latencies to assess whether empirical indicators of weak and strong satisficing take respondents, on average, less time to produce compared to indicators of optimized responses. Counter to what satisficing theory would predict, our findings show that the satisficing indicators were associated with significantly *longer* response latencies than optimized responses. One interpretation of these results is that a core assumption of the theory of survey-satisficing is invalid. Our less radical conclusion, however, is that variables commonly deployed by survey researchers as indicators of survey satisficing are problematic measures of the behaviour they are seeking to measure.