German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD
Working Paper Series

Working Paper No. 136

Paradata

Frauke Kreuter and Carolina Casas-Cordero

April 2010

Federal Ministry
of Education
and Research

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/2008 Heike Solga; since 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Contact: Council for Social and Economic Data (RatSWD) | Mohrenstraße 58 | 10117 Berlin | office@ratswd.de

Paradata

Frauke Kreuter and Carolina Casas-Cordero

University of Maryland (fkreuter[at]survey.umd.edu)

Abstract

Paradata – data about the process of survey production – have drawn increasing attention as the statistical world moves towards the implementation of quality metrics and measures to improve quality and save costs. This paper gives examples of various uses of paradata and discusses access to paradata as well as future developments.

Keywords: paradata, process data, responsive design, measurement error, nonresponse, adjustment

1. Introduction

During the last two decades, survey researchers have begun to use computer-assisted methods to collect social science data. This trend is most obvious in web surveys, but is equally present in telephone surveys that use automated call scheduling systems or mail surveys that take advantage of logs provided by postal services. All of these systems produce data about the survey process as by-product, which Mick Couper coined paradata in a presentation at the Joint Statistical Meeting in Dallas (Couper, 1998). Inspired by Couper's suggestions to use data automatically generated by computer-aided systems to evaluate survey quality, survey methodologists have since then broadened the concept of paradata to other aspects of the survey process and other modes of collection.

Data about the process have drawn increasing attention as the statistical world moves towards the implementation of quality metrics, measures to improve quality and save costs, and a framework of total survey error (Biemer and Caspar 1994; Lyberg et al. 1997; Aitken et al. 2004; Couper and Lyberg 2005). Both data users and data producers are now aware of the potential benefits of paradata, and this has been reflected in special interest and invited paper sessions at international conferences such as the International Workshop on Household Survey Nonresponse, bi-annual conferences of the European Survey Research Association (ESRA), annual conferences of the American Association of Public Opinion Research (AAPOR), Joint Statistical Meetings (JSM) and the Sessions of the International Statistical Institute (ISI), as well as the quality conferences co-organized by Eurostat.

2. Examples for paradata and their use

There is no standard definition in the literature of what constitutes paradata. Several papers attempt to systematize data that are not part of the actual interview (Scheuren 2000; Couper and Lyberg 2005; Scheuren 2005; O'Reilly 2009), but each of these papers varies slightly in terminology and in what is called paradata. Paradata was originally conceptualized as the data automatically generated as the by-product of the computer-assisted survey process (e.g., call record data and keystrokes), but the term has more recently been expanded to include information that may be recorded by interviewers (e.g., observations), or captured through additional systems (e.g., digital audio recording) (Couper 1998).

For this review we do not seek to provide a fixed definition of paradata. What is important in our opinion is the concept of data collected during and about the process. These data can be used to understand and improve the process (and subsequently the end result). Thus instead of

a definition, we give some examples of how paradata is currently being used around the world.

One set of data typically referred to as paradata are call records collected during the process of contacting a sample case. The time of contact (day and time), as well as the outcome of a call (non-contact, refusal, ineligible, interview, appointment, etc.) are almost always available on these call records (Heerwegh et al. 2007; Blom et al. forthcoming). These variables are either recorded by the interviewer (with PAPI or CAPI systems) or automatically, as is commonly the case for call schedulers in computer-aided telephone interviews (CATI). The recording of the date and time of a prior contact allows call schedulers to vary contact attempts with the hope of increasing the probability of a successful contact (Weeks et al. 1987; Kulka and Weeks 1998; Greenberg and Stokes 1990; Stokes and Greenberg 1990; Brick et al. 1996; Sangster and Meekins 2004; Wagner and Raghunathan 2007), and ideally to reduce the cost (Groves 1989; Triplett 2002; Murphy et al. 2003). Prominent examples of call record data collected in face-to-face surveys are the Contact History Instrument (CHI) implemented in surveys by the US Census Bureau (Bates 2003), or the standard contact forms that have been a required implementation since round one of the European Social Survey (Stoop et al. 2003). In some instances, call record data are used to guide decisions on responsive or two-phase sampling designs (Groves et al. 2003; Kennickell 2003; Groves and Heeringa 2006; Eckman and O'Muircheartaigh 2008), or to gain knowledge about optimal calling patterns in face-to-face surveys in general (Matsuo et al. 2006; Durrant et al. 2009). To our knowledge, there is so far only one survey, the US National Survey of Family Growth (Lepkowski et al. 2009), in which call record data from face-to-face surveys are used to drive centralized day-to-day field decisions similar to those in supervised call centers. For most surveys, face-to-face call record data are analyzed after the fact to assess interviewer efforts and compliance with pre-specified design requests (Billiet and Pleysier 2007; Lipps 2007; Koch et al. 2009).

Regardless of the mode of data collection, survey methodologists use call record data to study various aspects of survey participation. Call record data are available for both respondents and non-respondents in a survey and thus are prime candidates for the study of nonresponse bias, for example, through level-of-effort analyses, in which early respondents are compared to late responders assuming that later responders are more similar to nonresponders than early responders (Stinchcombe et al. 1981; Smith 1984; Schnell, 1998; Kennickell 1999; Chiu et al. 2001; Duhart et al. 2001; Lynn et al. 2002; Lynn 2003; Wang et al. 2005; Stoop 2005; Voogt and Saris 2005; Billiet et al. 2007); for a meta-analysis of the

results, see Olson (2010). With the goal of assessing net quality gains, researchers have used call record data to shed light on the relationship between nonresponse and measurement error (Green 1991; Yan et al. 2004; Olson 2006; Peytchev and Peytcheva 2007; Yu and Yan 2007).

A second set of data subsumed under the concept of paradata is also collected during the initial phase of establishing contact and convincing sample units to participate in the survey. These paradata are observations made by the interviewer. Like call record data, these interviewer observations are available on all sampled cases and thus suitable to inform survey design decisions (Copas and Farewell 1998; Lynn 2003; Groves et al. 2007) and assess nonresponse bias (Maitland et al. 2009). In recent face-to-face surveys, interviewers are charged with collecting observations of neighborhoods and housing unit characteristics in a number of surveys usually along the lines suggested by Campanelli et al. (1997), Groves and Couper (1998), or Lynn (2003). Examples are the US Health and Retirement Study, the US Study of Early Child Care, the US Survey of Consumer Finances, the US National Survey on Drug Use and Health, the British Election Study, the British Crime Survey, the British Social Attitudes Survey, the European Social Survey, and the Survey of Health, Ageing and Retirement in Europe. Some rather novel interviewer observations are those that are tailored to the survey topic and thus have higher potential to be useful for adaptive survey design decisions or nonresponse adjustment. Again a prime example is the National Survey of Family Growth, in which interviewers are asked to guess whether or not the sample person is currently in an active sexual relationship (with an opposite-sex partner), and whether or not children are present in the household Groves et al. (2007). Other sets of interviewer observations made at the doorstep are those capturing the interaction between interviewer and respondent and respondents' reasons for refusal (Campanelli et al. 1997; Bates and Piani 2005; Bates et al. 2008).

Both call record data and interviewer observations have the potential to enhance current nonresponse adjustments. Not only are they available for both respondents and nonrespondents, but ideally are they predictive of the sampled person's probability of responding to a survey and of the survey variables of interest. Over the years, survey methodologists have extensively researched and developed covariates of survey participation (Schnell, 2005; Groves and Couper 1998), many of which are now part of call record and contact data forms. The possibility of using call record data for nonresponse adjustment has been discussed for quite some time (Drew and Fuller 1980; Potthoff et al. 1993), and current papers demonstrate the relationship between information in call records and the probability of responding to a survey request (Beaumont 2005; Biemer and Wang 2007; Blom 2009;

Kreuter and Kohler 2009). Interviewer observations of variables close to the survey (such as the presence of children in a fertility survey) can complement call record data in response propensity models due to their likely stronger relationship to survey variables of interest (Kreuter et al. 2010), although difficult issues in modeling may arise when strong predictors of response are combined with strong predictors of survey outcome variables (Kreuter and Olson 2010).

In computer-aided surveys, a third set of paradata can be captured: audio-recordings of the interaction between interviewer and respondent. Researchers have suggested that vocal characteristics of the respondent and interviewer are in part responsible for successful recruitment attempts. Especially during telephone interviews, potential respondents have very little information about the interviewer, aside from how he/she sounds, speaks, and interacts when they decide whether or not to participate in a telephone interview (Groves et al. 2007; Best et al. 2009). Yet interviewers vary widely in how often their invitations lead to participation, suggesting that potential respondents may give considerable weight to interviewers' verbal attributes. Recordings and paradata derived from them are of interest, not only because they can shed light on survey participation, but also because they can be used to assess measurement errors on a question level (Jans 2010). Recordings become more common as digital storage becomes less expensive (Couper 2005; Thissen et al. 2007). However the post-processing of such recordings into usable paradata is a large task and has been undertaken in only a few methodological studies. Those studies make use of recent developments in the field of acoustical engineering and new software, which makes it possible for researchers to automatically process audio files and obtain objective data on voice characteristics such as disfluencies, pauses, interruptions, speech rate, and pitch (Jans 2010; Conrad et al. 2010).

In addition to audio-recordings, computer-assisted survey instruments facilitate the automated collection of paradata that can be used to assess measurement error at the question level. Most data collection software records the time used to complete a question, a set of questions, or the whole interview (response times), and capture key strokes, with which researchers can, for example, measure how often a respondent backed up and changed an answer and whether supplementary definitions are used (Couper 1998). All of these measures are available for computer-aided personal interviews (CAPI) and computer-aided telephone interviews (CATI) and Web surveys. Web surveys also differentiate between paradata that include characteristics of a respondent browser captured from server logs (server-side paradata) and respondent behavior captured by embedding JavaScript code into the instrument

(client-side paradata). Response times and key stroke measures have been used to study aspects of the response process (Bassili and Fletcher 1991; Kreuter, 2002; Heerwegh 2003; Kaminska and Billiet 2007; Yan and Tourangeau 2008; Couper et al., 2009; Lenzner et al. 2009; Peytchev 2009), to guide interventions in Web surveys (Conrad et al. 2009), evaluate interviewers (Couper et al. 1997; Mockovak and Powers 2008), and the review the performance of questions in pretests (Couper 2000; Stern 2008; Hicks et al. 2009).

Our list of examples is by no means complete, but it does give a flavor of the many uses of data auxiliary to the main data collection that contain information about the process with which the data are collected. There is in addition an entirely different usage of paradata beyond monitoring, managing, modeling, and improving the data collection process. Summary statistics of paradata are also used to describe the dataset as a whole: response rates (created out of recordings of the final status in call records) are examples of such survey-level statistics. While paradata contribute to such summary statistics, the summary statistics themselves are usually not referred to as paradata but called metadata instead (Couper and Lyberg 2005; Scheuren 2005).

Auxiliary data available on the case level that come from an entirely different source are also usually not considered paradata (i.e., administrative data, data from commercial lists, or data available on sampling frames). A more borderline case are separate surveys of the interviewers themselves (Siegel and Stimmel 2007). To the extent that information from interviewers can help to understand the survey process they can be viewed as paradata (like interviewer observations, for example). Metadata and auxiliary data also play increasing roles in monitoring and enhancing data quality. For some recent initiatives in using such auxiliary data see Smith (2007; 2009).

2.1 Databases and data access

Unlike survey data themselves and metadata about those surveys, paradata are usually not made publicly available for several reasons. For one, it is not common to release gross-sample data, i.e., data records that include all sampled units, both those that respond to the survey request and those that do not. Second, paradata are often not collected on the same unit of analysis as the survey data are, making the release of such data sets more complicated. Call record data are usually collected at each call attempt, which could easily generate up to fifty records for cases fielded in a telephone survey. Response times are collected at an item level and sometimes twice within one item (if the time to administer the item is measured separately from the time the respondent took to answer the question). Vocal properties of an

interviewer are recorded on a finer level and could generate several records even within the administration of a single item. Third, the format of these paradata varies a great deal by data collection agency and system: for example, outcome codes on call record data vary across agencies and modes of contact available to the interviewer (Blom et al. 2008). While the lack of standards for the collection and release of paradata is not a problem per se (except for making data preparation work more burdensome for analysts), it does require proper documentation, which is usually not covered by data collection grants. Fourth, for some of the paradata, there are open legal and ethical questions. Detailed observations of the neighborhood or housing unit might facilitate the de-identification of survey respondents. For Web surveys, Couper and Singer (2009) raise the question whether respondents should be informed about the capturing of client-side paradata in particular if they are used to understand or even control respondent behavior, and not just used for improvement of the design or performance of the instrument.

Some important surveys do release their paradata to the public; examples are contact protocol data from the European Social Survey, paradata from the U.S. National Health Interview Survey, and paradata from the American National Election Survey (the latter being available for secondary analysis upon request).

3. Future developments

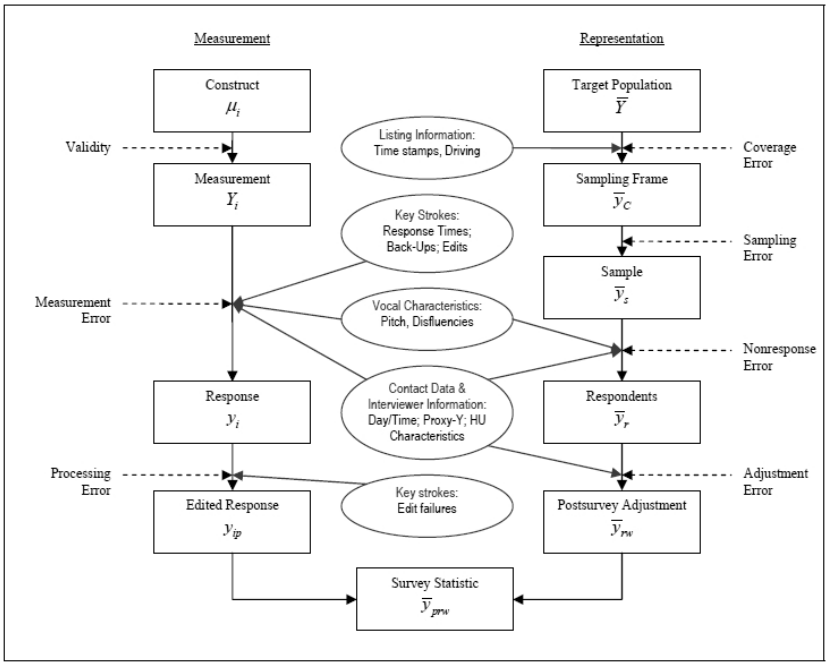
3.1 Data provision

As the previous section showed, the potential uses of paradata are wide-ranging. Survey methodologists have started to exploit paradata to guide intervention decisions during data collection and to provide opportunities for cost savings. To the extent that errors cannot be prevented, paradata also help us to detect errors after the fact (thus providing guidance for the next survey) and to model and adjust for them. So far a series of paradata have been used to assess or model measurement error, nonresponse error, and even the interaction of the two. Until now, very few paradata have been collected for other parts of the process. If we match the most commonly collected paradata to the various error sources in a total survey error framework (see Figure 1), we see that for several process steps in the generation of a survey statistics, no paradata are currently available. The systematic documentation of questionnaire development by Schnell et al. (2008) could lead to new paradata for the creation of measurement indicators.

From a quality monitoring and improvement perspective, a more structured approach towards the selection, measurement, and analysis of key process variables would be desirable (Morganstein and Marker 1997). Ideally survey researchers would specify a set of product characteristics and underlying processes associated with these characteristics, and then these processes would be checked by means of key process variables.

The virtue of paradata as a by-product of the survey process is that they come cheap to the data collector. If paradata are used systematically for process improvement and postprocess analyses, then their structure will probably change: variables will be added (e.g., new interviewer observations) and requests for standardization might turn out to conflict with existing collection systems. Paradata might then no longer be just a by-product, but a product with costs attached to it. It is up to survey methodologists to prove that paradata provide the cost control (or even cost savings) and performance increases that they have promised. Without the demonstration of repeated and successful use, survey methodologists will face difficulties in convincing data collection agencies to routinely collect such data.

Figure 1: Total Survey Error components and paradata for their assessment (modified graph from Groves et al. 2004)



One obstacle to demonstrating the usefulness of paradata is the quality of the data itself. While paradata might help to address some of the errors present in survey data, the data may suffer from measurement error, missing data, etc. Interviewers can erroneously record certain housing unit characteristics, can misjudge features about the respondents, or can fail to record a contact attempt altogether (Casas-Cordero 2010; Sinibaldi 2010; West 2010). For example,

it is possible that paradata are subject to high variation in the way the information is recorded by different interviewers (e.g., evaluation of the condition of the house relative to other houses in the area) or some interviewers may simply not place high priority on filling in the interviewer observation questionnaires because they are not paid for doing so. Some studies have shown high levels of missing data in interviewer observations, indicating a lack of data quality (Kreuter et al. 2007; Durrant et al. 2009). Such missings may occur, for example, if the interviewer does not have enough time or does not feel the need to fully record every contact attempt to the household. Likewise scripts embedded in Web surveys can fail to install properly and client-side data are not captured as intended, and recordings of interviewer administered surveys can be inaudible due to background noise or loose microphones (McGee and Gray 2007; Sala et al. 2008).

As long as these recording errors and missing data patterns are not systematic, they will reduce the effectiveness of paradata for process improvement and error modeling but should not threaten them altogether. If errors appear systematically (e.g., savvy users in Web surveys prevent scripts from capturing key strokes), resulting conclusions are threatened to be biased. Currently not enough is known to about the measurement error properties of paradata.

3.2 Data usage

As mentioned before, a key challenge to the use of paradata is their unusual data structure, with time-dependent observations on multiple levels collected through various modes with varying instruments. If we again take call record data as an example, the literature is still dominated by analyses using case-level aggregate statistics of call-level data (e.g., total number of contact attempts, total number of refusals), while some more recent examples take advantage of the multilevel structure by using survival models or multilevel discrete time event history models in predicting propensities to respond (Durrant and Steele 2009; Olson and Groves 2009; Wagner 2009).

Many methodological questions concerning how to make best use of paradata are still unsolved. In the estimation of response propensity models, we do not know yet if time should be modeled discretely as active day in the field or relative to the time since beginning of the field period. Nor is it clear how to best combine paradata into nonresponse propensity models with the aim of adjusting for survey nonresponse (Kreuter and Olson 2010). When dealing with response latencies, we do not yet know how best to handle unusually long response times, how best to model time dependency within the process of answering multiple

subsequent survey questions, etc. Closer collaboration among survey methodologists, statisticians, and econometric modelers could benefit the research in this area.

Methodologists who use paradata for management and monitoring are still experimenting with tools for displaying the constant flow of process information. A "dashboard" was developed at the Institute for Social Research in Michigan (Groves et al. 2008; Lepkowski et al. 2009) to provide survey managers and principal investigators with timely access to data, and tools to facilitate decision-making – but there is still room for improvement (Couper 2009). The use of process control charts has been proposed before (Deming 1986; Morganstein and Marker 1997; Couper and Lyberg 2005), but so far, no standard charts have been developed to monitor survey data collection. Increased access to paradata and in particular timely update of such data streams will increase the need for good tools to display and analyze paradata.

3.3 Data Access

To address the risk of de-identification of respondents, the paradata that pose this danger could be made available in research data centers where access and usage of data is monitored. Given the potential of certain paradata to improve nonresponse adjustment, an entirely new data retrieval system might be a worth considering. Given appropriate paradata, nonresponse adjustment can be tailored to individual analyses. Usually only one set of nonresponse adjustment weights is created and distributed with survey data. Growing nonresponse has made the assumption that a single adjustment strategy is sufficient for all statistics produced by a survey less tenable. A data retrieval system could be conceptualized that allows the on-demand creation of adjustment weights based on the planned analysis.

Public access to paradata also allows a post-hoc examination of the procedures followed by the data collection institutes. If survey organizations are aware that process information will become public, this might lead overall to a higher data collection standard. Obviously higher-quality work will come with a price. However, some survey organizations might not want to release paradata as it discloses information about their fieldwork procedures. If these procedures are considered to be proprietary, the disclosure could be seen as an impingement on their comparative advantage.

4. Discussion

Survey data collection is essentially a production process with a product. Surveys do not differ in this respect from other organizations that produce products or services and are concerned about their quality. Management strategies for such organizations have moved to what are called continuous quality improvement methods (Imai 1986; Deming 1986), in which measures of the process are monitored along the way so that error sources can be located and interventions planned (examples of such strategies are Total Quality Method, TQM, or Six Sigma). Several researchers have suggested the application of such strategies to the process of survey operations (Biemer and Caspar 1994; Morganstein and Marker 1997). Paradata as discussed here can play an important role in the application of such strategies. The European Statistical System has developed a handbook on improving quality through the analysis of paradata (Aitken et al. 2004), but the work is still not done, and individual surveys might do well to identify key process variables for their specific circumstances (Couper and Lyberg 2005).

Survey data collection faces major uncertainties in the planning stages. It is difficult to estimate the effectiveness of measures taken to establish contact with households, identify eligible persons, select a respondent, gain that person's cooperation, and complete the interview. Likewise, estimates of the cost implications of any of these steps are often difficult to make. Responsive designs (Groves and Heeringa 2006) seek to address this uncertainty by measuring the results of various survey design features, often experimentally, and then use these measurements to intervene in the field data collection process. This monitoring includes both the paradata as well as key survey estimates. To the extent that the paradata provide information about the risk of nonresponse bias, the responsive design is capable of reducing the risk of this bias. Much more effort is needed to manage the costs of alternative design features.

To increase the conditions for high-quality collection of paradata, a survey climate is necessary that allows for experimental manipulation within the field process. Pooling data across studies can also help to disentangle confounding elements; for this, some standard ization of paradata would be necessary (Blom et al. 2008). Panel data enjoy the luxury of repeated measures of observations. Researchers only recently started to explore the potential of paradata to examine attrition (Lepkowski and Couper 2002; Kreuter and Jäckle 2008) and measurement error in relation to interviewer characteristics (Jaeckle et al. 2009; Weinhardt and Kreuter 2009; Yan and Datta 2009).

Compared to other countries, data collection in Germany is not as “paradata-rich” as it could be. Since 1995, Schnell and his colleagues suggested the inclusion of contact protocol data for the gross sample to be a standard deliverable (Schnell et al. 1995). Very few surveys followed this suggestion. Furthermore, systems should be developed and put in place that allow data collection agencies to engage in data-driven interventions into the fieldwork process. For a single survey, the start-up costs might be too high and survey organizations might not see the need for such investments. If, however, the German social science data community as a whole demands paradata for process controls, investments in the respective systems might be economical. Investment is also needed into the development of new statistical tools and methods to help make sense of the vast amount of unstructured paradata generated by modern survey process. The standard analytic tools we use for survey data are not appropriate for much of the paradata we need to analyze. Here, too, collaboration throughout the social science data community would be a good first step.

References:

Aitken, A./Hörngren, J./Jones, N./Lewis, D. and Zilhão, M.J. (2004): Handbook on improving quality by analysis of process variables. Technical report, Eurostat.

Bassili, J.N. and Fletcher, J.F. (1991): Response-time measurement in survey research a method for CATI and a new look at nonattitudes. Public Opinion Quarterly 55 (3), 331-346.

Bates, N. (2003): Contact histories in personal visit surveys: The Survey of Income and Program Participation (SIPP) methods panel. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Bates, N./Dahlhamer, J. and Singer, E. (2008): Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. Journal of Official Statistics 24 (4), 591-612.

Bates, N. and Piani, A. (2005): Participation in the National Health Interview Survey: Exploring reasons for reluctance using contact history process data. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

Beaumont, J. (2005): On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. Survey Methodology 31 (2), 227-231.

Best, H./Bauer, G. and Steinkopf, L. (2009): Interviewer voice characteristics and productivity in telephone surveys. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Biemer, P. and Caspar, R. (1994): Continuous quality improvement for survey operations: Some general principles and applications. Journal of Official Statistics 10 (3), 307-326.

Biemer, P. and Wang, K. (2007): Using callback models to adjust for nonignorable nonresponse in face-to-face surveys. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Billiet, J./Philippens, M./Fitzgerald, R. and Stoop, I. (2007): Estimation of nonresponse bias in the European Social Survey: Using information from reluctant respondents. Journal of Official Statistics 23 (2), 135-162.

Billiet, J. and Pleysier, S. (2007): Response based quality assessment in the ESS – Round 2. An update for 26 countries. Technical report, Center for Sociological Research (CeSO), K.U. Leuven.

Blom, A. (2009): Nonresponse bias adjustments: What can process data contribute? Iser working paper 2009-21, Institute of Social research (ISER), Essex University.

Blom, A./Lynn, P. and Jäckle, A. (2008): Understanding cross-national differences in unit nonresponse: The role of contact data. Technical report, Institute for Social and Economic Research ISER.

Blom, A./Lynn, P. and Jäckle, A. (forthcoming): Understanding cross-national differences in unit nonresponse: The role of contact data. In: Harkness, J.A./Edwards, B./Braun, M./Johnson, T.P./Lyberg, L.E./Mohler, P.P./Pennell, B.-E. and Smith, T. (Eds.): Survey Methods in Multinational, Multiregional, and Multicultural Contexts. New York.

Brick, J.M./Allen, B./Cunningham, P. and Maklan, D. (1996): Outcomes of a calling protocol in a telephone survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Campanelli, P./Sturgis, P. and Purdon, S. (1997): Can you hear me knocking: An investigation into the impact of interviewers on survey response rates. Technical report, The Survey Methods Centre at SCPR, London.

Casas-Cordero, C. (2010): Testing competing neighborhood mechanisms influencing participation in household surveys. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, IL.

Chiu, P./Riddick, H. and Hardy, A. (2001): A comparison of characteristics between late/difficult and non-late/difficult interviews in the National Health Interview Survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Conrad, F./Broome, J./Benki, J./Groves, R.M./Kreuter, F. and Vannette, D. (2010): To agree or not to agree: Effects of spoken language on survey participation decisions. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, IL.

Conrad, F./Couper, M./Tourangeau, R./Galesic, M. and Yan, T. (2009): Interactive feedback can improve accuracy of responses in web surveys. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Copas, A. and Farewell, V. (1998): Dealing with non-ignorable nonresponse by using an 'enthusiasm-to-respond' variable. Journal of the Royal Statistical Society, Series A 161 (3), 385-396.

Couper, M. (1998): Measuring survey quality in a CASIC environment. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Couper, M. (2000): Usability evaluation of computer-assisted survey instruments. Social Science Computer Review 18 (4), 384-396.

Couper, M. (2005): Technology trends in survey data collection. Social Science Computer Review 23, 486-501.

Couper, M. (2009): Measurement error in objective and subjective interviewer observations. Modernisation of Statistics Production, Standardisation, Efficiency, Quality assurance and Customer Focus, 2-4 November 2009, Stockholm, Sweden.

Couper, M./Hansen, S. and Sadosky, S. (1997): Evaluating interviewer performance in a CAPI survey. In: Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N. and Trewin, D. (Eds.): Survey Measurement and Process Quality. New York.

Couper, M. and Lyberg, L. (2005): The use of paradata in survey research. In: Proceedings of the 55th Session of the International Statistical Institute, Sydney, Australia.

Couper, M. and Singer, E. (2009): Ethical considerations in the use of paradata in web surveys. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Couper, M./Tourangeau, R. and Marvin, T. (2009): Taking the audio out of Audio-CASI. Public Opinion Quarterly 73 (2), 281-303.

Deming, W. (1986): Out of the Crisis. Cambridge.

Drew, J. and Fuller, W. (1980): Modeling nonresponse in surveys with callbacks. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Duhart, D./Bates, N./Williams, B./Diffendal, G. and Chiu, P. (2001): Are late/difficult cases in demographic survey interviews worth the effort? A review of several federal surveys. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

Durrant, G./D'Arrigo, J.and Steele, F. (2009): Using field process data to predict best times of contact conditioning on household and interviewer influences. Technical report, Southampton Statistical Science Research Institute.

Durrant, G. and Steele, F. (2009): Multilevel modelling of refusal and noncontact in household surveys: Evidence from six UK government surveys. Journal of the Royal Statistical Society, Series A 172 (2), 361-381.

Eckman, S. and O'Muircheartaigh, C. (2008): Optimal subsampling strategies in the General Social Survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Eifler, S./Thume, S. and Schnell, R. (2009): Unterschiede zwischen subjektiven und objektiven Messungen von Zeichen öffentlicher Unordnung. In: Weichbold, M./Bacher, J. and Wolf, Ch. (Eds.): Umfrageforschung. Herausforderungen und Grenzen. Wiesbaden. (Sonderheft 9 der Österreichischen Zeitschrift für Soziologie).

Green, K.E. (1991): Reluctant respondents: Differences between early, late and nonrespondents to a mail survey. Journal of Experimental Education 56, 268-276.

Greenberg, B. and Stokes, S. (1990): Developing an optimal call scheduling strategy for a telephone survey. Journal of Official Statistics 6 (4), 421-435.

Groves, R.M. (1989): Survey Errors and Survey Costs. New York.

Groves, R.M. and Couper, M. (1998): Nonresponse in Household Interview Surveys. New York.

Groves, R.M./Fowler, F./Couper, M. and Lepkowski, J. (2004): Survey Methodology. New York.

Groves, R.M. and Heeringa, S. (2006): Responsive design for household surveys: Tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society, Series A 169 (3), 439-457.

Groves, R.M./Wagner, J. and Peytcheva, E. (2007): Use of interviewer judgments about attributes of selected respondents in post-survey adjustment for unit nonresponse: An illustration with the National Survey of Family Growth. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Groves, R.M./Kirgis, N./Peytcheva, E./Wagner, J./Axinn, B. and Mosher, W. (2008): Responsive design for household surveys: Illustration of management interventions based on survey paradata. NCRM Research Methods Festival, St Catherine's College, Oxford, UK.

Groves, R.M./Van Hoewyk, J./Benson, F./Schultz, P./Maher, P./Hoelter, L./Mosher, W./Abma, J. and Chandra, A. (2003): Using process data from computer-assisted face to face surveys to help make survey management decisions. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Nashville, TN.

Heerwegh, D. (2003): Explaining response latencies and changing answers using client-side paradata from a web survey. Social Science Computer Review 21 (3), 360-373.

Heerwegh, D./Abts, K. and Loosveldt, G. (2007): Minimizing survey refusal and noncontact rates: Do our efforts pay off? Survey Research Methods 1 (1), 3-10.

Herget, D./Biemer, P./Morton, J. and Sand, K. (2005): Computer audio recorded interviewing (CARI): Additional feasibility efforts of monitoring field interview performance. Paper presented at the U.S. Federal Conference on Statistical Method.

Hicks, W./Edwards, B./Tourangeau, K./Branden, L./Kistler, D./McBride, B./Harris-Kojetin, L. and Moss, A. (2009): A system approach for using CARI in pretesting, evaluation and training. Paper presented at the FedCasic Conference, Delray Beach, FL.

Imai, M. (1986): Kaizen: The Key to Japan's Competitive Success. New York.

Jaeckle, A./Sinibaldi, J./Tipping, S./Lynn, P. and Nicolaas, G. (2009): Interviewer characteristics, their behaviours, and survey outcomes.

Jans, M. (2010): Verbal Paradata and Survey Error: Respondent Speech, Voice, and Question-Answering Behavior Can Predict Income Item Nonresponse. Ph. D. thesis, University of Michigan, United States.

Kaminska, O. and Billiet, J. (2007): Satisficing for reluctant respondents in a cross-national context. Paper presented at the European Survey Research Association (ESRA) Conference, Prague, Czech Republic.

Kennickell, A. (1999): What do the 'late' cases tell us? Evidence from the 1998 Survey of Consumer Finances. Paper presented at the International Conference on Survey Nonresponse, Portland, OR.

Kennickell, A. (2003): Reordering the darkness: Application of effort and unit nonrepsonse in the Survey of Consumer Finances. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Koch, A./Blom, A./Stoop, I. and Kappelhof, J. (2009): Data collection quality assurance in cross-national surveys: The example of the ESS. Methoden, Daten, Analysen 3 (2), 219-247.

Kreuter, F. (2002): Kriminalitätsfurcht: Messung und methodische Probleme. Opladen.

Kreuter, F. and Jäckle, A. (2008): Are contact protocol data informative for non-response bias in panel studies? A case study form the Northern Ireland subset of the British Household Panel Survey. Panel Survey Methods Workshop, University of Essex, Colchester, UK.

Kreuter, F. and Kohler, U. (2009): Analyzing contact sequences in call record data. Potential and limitation of sequence indicators for nonresponse adjustment in the European Social Survey. Journal of Official Statistics 25 (2), 203-226.

Kreuter, F./Lemay, M. and Casas-Cordero, C. (2007): Using proxy measures of survey outcomes in post-survey adjustments: Examples from the European Social Survey (ESS). In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Kreuter, F. and Olson, K. (2010): Multiple auxiliary variables in nonresponse adjustment. [Manuscript under review].

Kreuter, F./Olson, K./Wagner, J./Yan, T./Ezzati-Rice, T./Casas-Cordero, C./Lemay, M./Peytchev, A./Groves, R.M. and Raghunathan, T. (2010): Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: Examples from multiple surveys. Journal of the Royal Statistical Society, Series A.

Kulka, R. and Weeks, M. (1998): Toward the development of optimal calling protocols for telephone surveys: A conditional probabilities approach. Journal of Official Statistics 4 (4), 319-332.

Lenzner, T./Kaczmirek, L. and Lenzner, A. (2009): Cognitive burden of survey questions and response times: A psycholinguistic experiment. Applied Cognitive Psychology.

Lepkowski, J. and Couper, M. (2002): Nonresponse in longitudinal household surveys. In: Groves, R.M./Dillman, D./Eltinge, J. and Little, R. (Eds.): Survey Nonresponse. New York.
Lepkowski, J./Groves, R.M./Axinn, W./Kirgis, N. and Mosher, W. (2009): Use of paradata to manage a field data collection. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Lipps, O. (2007): Cooperation in centralised CATI panels – A contact based multilevel analysis to examine interviewer, respondent, and contact effects. Paper presented at the European Survey Research Association (ESRA) Conference, Prague, Czech Republic.
Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N. and Trewin, D. (1997): Survey Measurement and Process Quality. New York.
Lynn, P. (2003): PEDAKSI: Methodology for collecting data about survey non-respondents. Quality and Quantity 37 (3), 239-261.
Lynn, P./Clarke, P./Martin, J. and Sturgis, P. (2002): The effects of extended interviewer efforts on nonresponse bias. In: Dillman, D./Eltinge, J./Groves, R.M. and Little, R. (Eds.): Survey nonresponse. New York.
Maitland, A./Casas-Cordero, C. and Kreuter, F. (2009): An evaluation of nonresponse bias using paradata from a health survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Matsuo, H./Loosveldt, G. and Billiet, J. (2006): The history of the contact procedure and survey cooperation – Applying demographic methods to European Social Survey contact forms round 2 in Belgium. Louvain-la-Neuve, Belgium. Paper presented at the Quetelet Conference.
McGee, A. and Gray, M. (2007): Designing and using a behaviour code frame to assess multiple styles of survey items. Technical report, National Centre for Social Research (NatCen), London, UK.
Mockovak, W. and Powers, R. (2008): The use of paradata for evaluating interviewer training and performance. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Morganstein, D. and Marker, D. (1997): Continuous quality improvement in statistical agencies. In: Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N. and Trewin, D. (Eds.): Survey Measurement and Process Quality. New York.
Murphy, W./O'Muircheartaigh, C./Emmons, C./Pedlow, S. and Harter, R. (2003): Optimizing call strategies in RDD: Differential nonresponse bias and costs in REACH 2010. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Olson, K. (2006): Survey participation, nonresponse bias, measurement error bias, and total bias. Public Opinion Quarterly 70 (5), 737-758.
Olson, K. (2010): When do nonresponse follow-ups improve or reduce data quality? A synthesis of the existing literature. [Manuscript].
Olson, K. and Groves, R.M. (2009): The lifecycle of response propensities in fertility and family demography surveys. Paper presented at the annual meeting of the Population Association of America, Detroit, MI.
O'Reilly, J. (2009): Paradata and Blaise: A review of recent applications and research. Paper presented at the International Blaise Users Conference (IBUC), Latvia.
Peytchev, A. (2009): Survey breakoff. Public Opinion Quarterly 73 (1), 74-97.
Peytchev, A. and Peytcheva, E. (2007): Relationship between measurement error and unit nonresponse in household surveys: An approach in the absence of validation data. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Potthoff, R.,/Manton, K. and Woodbury, M. (1993): Correcting for nonavailability bias in surveys by weighting based on number of callbacks. Journal of the American Statistical Association 88 (424), 1197-1207.
Sala, E./Uhrig, S. and Lynn, P. (2008): The development and implementation of a coding scheme to analyse interview dynamics in the British Household Panel Survey. Technical report, Institute for Social and Economic Research ISER, University of Essex.
Sangster, R. and Meekins, B. (2004): Modeling the likelihood of interviews and refusals: Using call history data to improve efficiency of effort in a national RDD survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Scheuren, F. (2000): Macro and micro paradata for survey assessment. [Manuscript].
Scheuren, F. (2005): Paradata from concept to completion. In: Proceedings of the Statistics Canada Symposium. Methodological Challenges for Future Information Needs.
Schnell, R. (1998): Besuchs- und Berichtsverhalten der Interviewer. In: Statistisches Bundesamt (Ed.): Interviewereinsatz und –qualifikation. Stuttgart.
Schnell, R. (2005): Nonresponse in Bevölkerungsumfragen. Opladen. www.ub.uni-konstanz.de/kops/volltexte/2008/5614. [Last visited: 03/10/2010].
Schnell, R./Hill, P. and Esser, E. (1995): Methoden der empirischen Sozialforschung. 5th Edition [latest edition 2008]. Munic.
Schnell, R./Krause, J./Stempfhuber, M./Zwingenberger, A. and Hopt, O. (2008): Softwarewerkzeuge zur Dokumentation der Fragebogenentwicklung – QDDS – 2. Technical report, Endbericht des Projekts an die DFG.
Siegel, N. and Stimmel, S. (2007): SOEP-Interviewerbefragung 2006. Methodenbericht. Technical report, Munich.
Sinibaldi, J. (2010): Measurement error in objective and subjective interviewer observations. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, MI.
Smith, T. (1984): Estimating nonresponse bias with temporary refusals. Sociological Perspectives 27 (4), 473-489.
Smith, T. (2007): The Multi-level Integrated Database Approach (MIDA) for improving response rates: Adjusting for nonresponse error, and contextualizing analysis. Paper presented at the European Survey Research Association (ESRA) Conference, Prague, Czeck Republic.
Smith, T. (2009): The Multi-level Integrated Database Approach for detecting and adjusting for nonresponse bias. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.
Stern, M. (2008): The use of client-side paradata in analyzing the effects of visual layout on changing responses in web surveys. Field Methods 20 (4), 377-398.
Stinchcombe, A./Jones, C. and Sheatsley, P. (1981): Nonresponse bias for attitude questions. Public Opinion Quarterly 45, 359-375.

Stokes, S. and Greenberg, B. (1990): A priority system to improve callback success in telephone surveys. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Stoop, I. (2005): The Hunt for the Last Respondent. The Hague.
Stoop, I./Devacht, S./Billiet, J./Loosveldt, G. and Philippens, M. (2003): The development of a uniform contact description form in the ESS. Paper presented at the International Workshop for Household Survey Nonresponse, Leuven, Belgium.
Thissen, R./Sattaluri, S./McFarlane, E. and Biemer, P.P. (2007): Evolution of audio recording in field surveys. Paper presented at The American Association for Public Opinion Research (AAPOR) 62th Annual Conference.
Triplett, T. (2002): What is gained from additional call attempts and refusal conversion and what are the cost implications? Technical report, The Urban Institute, Washington DC.
Voogt, R. and Saris, W. (2005): Mixed mode designs: Finding the balance between nonresponse bias and mode effects. Journal of Official Statistics 21, 367-387.
Wagner, J. (2009): Adaptive contact strategies in a telephone survey. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.
Wagner, J. and Raghunathan, T. (2007): Bayesian approaches to sequential selection of survey design protocols. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Wang, K./Murphy, J./Baxter, R. and Aldworth, J. (2005): Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.
Weeks, M./Kulka, R. and Pierson, S. (1987): Optimal call scheduling for a telephone survey. Public Opinion Quarterly 51, 540-549.
Weinhardt, M. and Kreuter, F. (2009): The different roles of interviewers: How does interviewer personality affect respondents' survey participation and response behavior?
West, B. (2010): An examination of the quality and utility of interviewer estimates of household characteristics in the National Survey of Family Growth. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, MI.
Yan, T. and Datta, R. (2009): Estimating the value of project-specific and respondent specific interviewer experience: Evidence from longitudinal and repeated cross-section surveys.
Yan, T. and Tourangeau, R. (2008): Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. Applied Cognitive Psychology 22 (1), 51-68.
Yan, T./Tourangeau, R. and Arens, Z. (2004): When less is more: Are reluctant respondents poor reporters? In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Yu, M. and Yan, T. (2007): Are nonrespondents necessarily bad reporters? Using imputation techniques to investigate measurement error of nonrespondents in an alumni survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.