Survey methodology: international developments

Kreuter, Frauke

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
SSG Sozialwissenschaften, USB Köln

German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD
Working Paper Series

Working Paper No. 59

Survey Methodology: International Developments

Frauke Kreuter

March 2009

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.

Survey Methodology: International Developments

Frauke Kreuter

Abstract

Falling response rates and the advancement of technology have shaped the discussion in survey methodology in the last few years. Both led to a notable change in data collection efforts. Survey organizations try to create adaptive recruitment and survey designs and increased the collection of non-survey data for sampled cases. While the first strategy is an attempt to increase response rates and to save cost, the latter is part of efforts to reduce possible bias and response burden of those interviewed. To successfully implement adaptive designs and alternative data collection efforts researchers need to understand error properties of mixed-mode and multiple-frame surveys. Randomized experiments might be needed to gain that knowledge. In addition close collaboration between survey organizations and researchers is needed, including the possibility and willingness to shared data between those organizations. Expanding options for graduate and post-graduate education in survey methodology might help to increase the possibility for high quality surveys.

Keywords: Survey Methodology, Responsive Design, Paradata

Falling response rates (Schnell 1997; Groves and Couper 1998; Atrostic 2001; De Leeuw and De Heer 2002) and the advancement of technology (Couper 2005) have shaped the discussion in survey methodology in the last few years. This section will highlight some of the developments that have resulted from these two trends and the increasing difficulty of conducting surveys in a way that was common and prominent during the 70s, 80s and 90s. It is impossible to capture all of the changes in survey practice during that time frame. Instead, I will address a few developments that are prominently discussed within the survey methodology community and are not captured in the other chapters of this volume.

The developments discussed here share an increased flexibility in data collection efforts while, at the same time, design changes are implemented in a controlled or even randomized way to evaluate their effects on the individual error sources. There is less of a streamlined, recipe-style approach to data collection; we see a move towards techniques that allow for adaptation either to recruit respondents or during the response process to lower respondent burden and interviewer effects. Unlike in Germany, the data infrastructure in the U.S. and U.K. allow for such flexibilities with survey organizations closely tied to scientists at Universities (e.g. University of Michigan), or survey research organizations that act as primary investigators (e.g. NORC for the General Social Survey). In both countries, most of the data collection agencies, used for social science research, are organizations that specialize in surveys for research projects. The companies therefore tend to have incentives to invest in developing the expertise to conduct high quality surveys.

This chapter will begin with a discussion of response rates as a quality indicator for surveys and summarize the current discussion of alternative indicators to response rates. I will then highlight recent developments within survey operations. Many of these developments are reactions to falling response rates and increased worries about nonresponse bias; others are motivated more broadly through the context of total survey error (Groves et al. 2004) or a reaction to changes in technology. The main question behind all these developments is “How can we ensure high quality data collection in changing environments and increase quality in existing ones?”

1. Response Rates and other Survey Quality Indicators

For years survey methodologists as well as the public have focused on response rates as quality indicators (Groves et al. 2008). This focus has changed in recent years. For one, even surveys with traditionally high response rates have fallen below expectations. In addition

empirical evidence over the last decade has increasingly demonstrated that nonresponse rates are poor indicators of nonresponse bias for single survey estimates (Keeter et al. 2000; Curtin et al. 2000; Groves 2006). The shift in focus, away from nonresponse rates towards bias is visible in several ways. It is visible for example in the requirement of a detailed plan to evaluate nonresponse bias before the U.S. Office of Management and Budget approves data collections sponsored by federal statistical agencies.[1] It is also visible in research initiatives to develop alternative indicators for survey quality (Groves et al. 2008).

Those alternative indicators can be grouped into two sets (Groves et al. 2008) – single indicators at the survey level (which is similar to the current use of the response rate), and individual indicators at the estimate level. Among the set of single indicators are *variance functions of nonresponse weights* (e.g., coefficients of variation of nonresponse weights), *variance functions of post-stratification weights* (e.g., coefficients of variation of poststratification weights), *variance functions of response rates* on subgroups defined for all sample cases (both respondents and nonrespondents), *goodness of fit statistics on propensity models,* and *R-indexes* (Shouten and Cobben 2007), which are model-based equivalents of the above. Researchers from the Netherlands, UK, Belgium, Norway and Slovenia formed a joint project (RISQ) to develop and study such R-indexes.

The second set of indicators is produced on a survey estimate level. It is evident that nonresponse bias is item specific (Groves and Peytcheva 2008) and thus estimate-level indicators would have the soundest theoretical basis. Examples of estimate specific indicators are: *comparisons of respondents and nonrespondents on auxiliary variables; correlation between post-survey nonresponse adjustment weights and the analysis variable of interest (y) measured on the respondent cases; variation of means of a survey variable y within deciles of the survey weights;* and *fraction of missing information on y.* The latter is based on the ratio of between-imputation variance of an estimate and the total variance of an estimate based on imputing values for all the nonrespondent cases in a sample (Little and Rubin 2002; Wagner 2008; Andrige and Little 2008).

All of these attempts rely heavily on the availability of auxiliary variables, such as enriched sampling frames, interviewer observations, or other paradata correlated with the

1 All data collection conducted or sponsored by the U.S. federal statistical agencies has to be approved by the Office of Management and Budget (OMB), which ensures that performance standards developed by the Interagency Council on Statistical Policy (ICSP) are met (Graham 2006). Conducting or sponsoring is defined here as any information that the agency collects using (1) its own staff and resources, or (2) another agency or entity through a contract or cooperative agreement. The approval by OMB is not just an attempt to reduce burden on the respondents (see Paperwork Reduction Act) but to ensure "that the concepts that are being measured be well known and understood, and shown to be reliable and valid" (Graham 2006). OMB applications require information from the data collection agency on questionnaire design procedures, field tests of alternative versions of their measures, reinterviews with subsamples of respondents, and the like. Pretests and pilot studies are encouraged, and the OMB guidelines spell out how those can be conducted. No criteria are specified to quantify potential measurement error. The development of a plan to evaluate nonresponse bias is required only in cases where projected unit response rate falls below 80 percent.

survey variables of interest. Thus we cannot revise our survey quality indicators without also changing survey operations.

Survey operations, the procedures of data collection, are themselves subject to quality assessment and quality indicators. O'Muircheartaigh and Heeringa (2008) presented a set of criteria at the 3MC conference in Berlin, another example for quality assessment of survey operations are the OMB guidelines.[2] Independent of those guidelines, there are a couple of recent developments in survey operations that are informative for the German data collection context.

2. Survey Operations

While survey methodologists and statisticians are aware of the fact that response rates are a poor indicator of nonresponse error (Keeter et al. 2000; Groves 2006) and are even less suitable as indicator of the overall survey quality (Groves et al. 2004), the drop in response rates has nevertheless engaged survey researchers in rethinking current practices. The increasing difficulty to gain cooperation has heightened the awareness of potential biases in surveys and created a need to evaluate survey procedures given the threat of losing precision through decreasing sample sizes. Changes in field work procedures require cost-quality trade-off decisions.

Surveys conducted with a *responsive design* use *paradata* to carryout those cost-quality trade-off decisions during the field work period. Such paradata are not only used to inform decision making during field operations but are increasingly seen as tools to evaluate measurement error or conduct post-survey bias adjustments. *Multiple-mode* surveys are often a response to cost-quality trade-off analyses prior to the start of the survey, but they are also a reaction to coverage problems that arise when mode-specific frames do not cover the entire population. An extreme form of multiple-mode surveys are those where the respondent recruitment is separated from the actual data capture. The most prominent examples are access panels or opt-in polls (discussed in other chapters of this volume and therefore omitted here).

Responsive Design

Survey organizations have used subsampling and two-phase designs for a long time. However, such design decisions were often only informed by estimates of a current response

2 http://www.whitehouse.gov/omb/inforeg/statpolicy.html#pr.

rate and qualitative information from field supervisors. Those efforts were also hampered by the inability to reach every sample unit in the subsample, and with that the statistical properties in the two-phase design are not necessarily unbiased. The second-phase response rates are related to the efforts used in the second phase, as well as first-phase design characteristics such as nonresponse, mode of data collection, and incentive structure. During the last decade survey organizations within the U.S. and some European countries started to systematically base design decisions on quantitative information gathered during early phases of the fieldwork period. The most prominent and detailed published example is given by the Social Research Center at Institute for Social Research of the University of Michigan and is labeled as responsive design (Groves and Heeringa 2006). These responsive designs are characterized by four features. First, a set of design characteristics is identified that potentially effect cost and errors of survey estimates. Second, this set of indicators is monitored through the initial stages of the data collection. Third, in subsequent phases of the data collection the features of the survey are altered based on cost-error trade-off decision rules. Finally, data from the separate phases are combined into a single estimator. An example of data collected are interviewer hours that are spent calling on sample households, driving to sample areas, conversing with household members and interviewing sample people.

One key element of such a responsive design is the ability to track key estimates as a function of estimated response propensities (conditioned on a design protocol). If survey variables can be identified that are highly correlated with the response propensity, and it can be seen that point estimates of such key variables are no longer affected by extending the field period, then one can conclude that the first phase of a survey (with a given protocol) has reached its phase capacity and a switch in recruitment protocol is advisable. Using non-contact error as an example, one can expect that a given recruitment protocol has reached its capacity if the percentage of households with access impediments stabilizes over repeated application of the recruitment protocol (e.g. repeated call-backs). Applying this method, Groves and Heeringa (2006) concluded for the National Survey of Family Growth cycle-6 field period that 10–14 calls produced stable cumulative estimates on the vast majority of the key estimates. A necessary condition for tracking key survey estimates concurrently is the possibility and willingness of interviewers not only to record respondent data and paradata electronically but also to submit the data in a timely manner to the survey managers. In the case of NSFG, the submissions occurred every evening (Wagner 2008).

Paradata

Paradata (data about the process of data collection) were already mentioned in the previous section as an important tool to guide fieldwork decisions. Increasingly paradata are also used as tools for survey nonresponse adjustment, and for detection and modeling of measurement error. The latter is already more common in Web surveys, where key-stroke files are readily available due to the nature of the task. Even face-to-face surveys now include the capability for electronically capturing survey-process data. Examples are key-stroke files obtained from computer-assisted personal interview (CAPI) or audio computer-assisted self interview (Audio-CASI) surveys (Couper et al. 2008), or through digital recordings of the (partial) interviews.

Paradata of potential use for nonresponse adjustments are collected in conjunction with household listings and when contact attempts are made to sample units. Recently, the U.S. Census Bureau implemented an automated system for collecting contact histories for CAPI surveys (Bates et al. 2006). Other governments have started using similar procedures. For example, the Research Center of the Flemish Government (Belgium) began to use contact forms in their surveys based on the work of Campanelli et al. (1997). The time of contact (day and time), the data collection method (in-person or by telephone), and other information is recorded for each contact attempt with each sample unit (Heerwegh et al. 2007). A standard contact form has also been implemented since 2002 (round one) of the European Social Survey, and contact data were recently released publicly by the U.S. National Center for Health Statistics (NCHS) for the 2006 National Health Interview Survey. Thus contact protocol data increasingly are available for each sample unit which makes those data an attractive source for nonresponse-adjustment variables. Other large survey projects collecting observations of neighborhoods and housing unit characteristics include the 2006 Health and Retirement Study (HRS), Phase IV of the Study of Early Child Care (SECC), the Survey of Consumer Finances (SCF), the National Survey on Drug Use and Health (Cunningham et al 2005), the British Election Study (BES), the British Crime Survey (BCS), the British Social Attitudes Survey (BSA), and the Survey of Health, Ageing and Retirement in Europe (SHARE).

Inspired by Groves and Couper (1998), some researchers linked interviewer observations to the likelihood of response. Copas and Farewall (1998) successfully used the interviewer-assessed interest of sample members about participating in the British National Survey of Sexual Attitudes and Lifestyle as a predictor for the likelihood of response. Lynn (2002) demonstrated how observations on multi-unit structures and door intercoms predict the

amount of effort that is required to contact sample households in the British Crime Survey. Bates et al. (2006) used contact information from the 2005 National Health Interview Survey (NHIS) to predict survey participation. They examined the effect of various respondent questions, concerns, and reasons for reluctance recorded by interviewers on the survey response. For the U.S. National Survey of Family Growth where Groves and Heeringa (2006) used a series of process and auxiliary variables to predict a screening and interview propensity for each active case. Those variables can be groups into observations on each interviewer, observations on segments made by people listing, observations about each listed housing unit made by people listing addresses, observations on each call to the, observations on each call yielding a contact with a household member, observations for each interviewer day (the number of hours spent travelling to the segment, the number of hours spent on administrative activities, the number of hours spent attempting screening interviews, the number of hours spent attempting main interviews). The expected screening and interview propensities were summed over all cases within a sample segment and grouped into propensity strata. The propensity strata were used by supervisors to direct the work of interviewers. Propensity models using call-record paradata were also estimated for the Wisconsin Divorce Study (Olson 2007) and the U.S. Current Population Survey (Fricker 2007). Both, Olson (2007) and Fricker (2007) then examined measurement error as a function of response propensity. Lately, more studies have tried to establish a relationship between paradata collected during the contact process (or as interviewer observations) and key survey variables (Schnell and Kreuter 2000; Asef and Riede 2006; Olson and Peytchev 2007; Groves et al. 2007; Yan and Raghunathan 2007; Kreuter et al. 2007). A systematic evaluation of the quality of such paradata, however, is very limited. For example, measurement error properties of these data, collected either through interviewer observation or through digital recordings of timing or speech are currently being studied by Casas-Cordero (2008) and Jans (2008).

Auxiliary variables and alternative frames

Next to paradata there is a second set of data sources that is now increasingly of interest to survey designers, that of commercial mass mailing vendors. These lists are of interest in the creation of sampling frames, to enhance survey information and to evaluate nonresponse bias.

In face-to-face surveys in the U.S. two methods of in-field housing unit listing are most common. Traditional listing in which listers are provided with maps showing the selected area and an estimate of the number of housing units they will find. Dependent listing in which listers are given sheets preprinted with addresses believed to lie inside the selected area.

Those addresses come either from a previous listing or from a commercial vendor. Listers travel around the segment and make corrections to the list to match what they see in the field. The latter appears to be less expensive (O'Muircheartaigh et al. 2003). There is a third method of creating a housing unit frame, which involves procuring lists of residential addresses from a commercial vendor and identifying those that fall inside the selected areas, geo-coding will be used instead of actual listings. The coverage properties of such frames still under study (Iannacchione et al. 2003; O'Muircheartaigh et al. 2006; Dohrmann et al. 2007; O'Muircheartaigh et al. 2007; Eckmann 2008). Survey research organizations are currently exploring the U.S. Postal-Service delivery sequence files to replace traditionally used PSUs (Census blocks) with Zip-Codes. While this last development is U.S. specific, it is nevertheless of interest as it provides the possibility to stratify with rich datasets, or to pre-inform interviewers about potential residents and their characteristics. This information can be used for tailored designs. In Germany, dependent listing and enhanced stratification was already used for the IAB-PASS study (Schnell 2007).

Multiple-modes

Several U.S. federal statistical agencies have explored the use of mixed-mode surveys. The two main reasons mixed-mode studies are usually considered are cost and response rates. There are three prominent types of multiple-mode studies: modes are administered in sequence, modes are implemented simultaneously, and a primary mode is supplemented with a secondary mode (DeLeeuw 2005).

The American Community Survey (ACS), which replaced the Census long form, is an example of a sequential application of modes. Respondents are first contacted by mail, nonrespondents to the mail survey are contacted on the phone (if telephone numbers can be obtained) and finally in-person follow-ups are made to a sample of addresses that remain uninterviewed. Parallel to the primary data collection a method sample is available to examine various error sources (Griffin 2008). The Bureau of Labor Statistics (BLS) is currently using multiple modes for the Current Employment Statistics (CES) program. Firms are initiated into the survey via computer-assisted telephone interview(CATI), kept on CATI for several month, and are then rolled over to touchtone data entry, the internet, fax etc.[3] Experiments are undertaken to evaluate measurement error separately from nonresponse error for each of these modes (Mockovak 2008). The National Survey of Family Growth has as its primary mode CAPI, but sensitive information (e.g., number of abortions) is collected through Audio-CASI.

3 http://www.bls.gov/web/cestn1.htm.

With the attraction of responsive design and the acknowledgement of imperfect sampling frames, mixed-mode surveys are attractive. Research is underway to explore the interaction between nonresponse and measurement error for these designs (Voogt and Saris 2005; Krosnick 2005). The European Social Survey program just launched a special mixed-mode design in four countries to examine appropriate ways of tailoring data collection strategies, and to disentangle mode effects into elements arising from measurement, coverage, and sample selection. Another large scale study within Europe that experiments with mixed-modes is the UK Household Longitudinal Study (UKHLS) under supervision of the Institute for Social and Economic Research. On the administrative side, the Social Research Center at the University of Michigan is currently constructing a new sample management system that will allow more efficient ways to carry out mixed-mode surveys (Axxin et al. 2008). The new system will manage sample across data collection modes (F2F, Telephone, Web and supplementary data modes such as biomarkers, soil samples etc.) and will allow easy transfer of sample between modes and interviewers (e.g., between CAPI and centralized CATI).

Reduction of response burden

Another development related to measurement error can be seen most recently in large scale surveys. Researchers at the Bureau of Labor Statistics (BLS) are investigating survey re-design approaches to reduce respondent burden in the Consumer Expenditure Survey (Gonzalez and Eltinge 2008). One proposed method is multiple matrix sampling, a technique for dividing a questionnaire into subsets of questions and then administering them to random subsamples of the initial sample. Matrix sampling has, for a long time, been used in large-scale educational testing. This method is growing in popularity for other types of surveys (Couper et al. 2008) where respondent burden is an increasing concern. Another method from educational testing that is currently under exploration is adaptive testing. Most applications of this method are currently tested in health surveys but survey issues regarding context effects arise (Kenny-McCoullough 2008).

Interviewer

All of the above mentioned developments have one feature in common. They alter and extend the task interviewers have to perform. In the past there was already a tension between the dual role of interviewers, as they are adaptive and flexible when recruiting respondents into the sample (Groves and McGonagle 2001; Maynard and Schaeffer 2002) and on the other hand neutral question delivery entities with ideally as little individuality as possible to reduce

interviewer effects (Schnell and Kreuter 2005). Now, the number of tasks required from one interviewer is even higher, as they span observations, book-keeping, handling technology, explaining technology, switching between different questionnaire flows, and the like. With this increased burden/expectations on the interviewer, a more careful look at interviewer performance seems necessary. Survey organizations (NORC in the U.S., NatCen and ONS in the U.K.) have already started to analyze interviewer performance across various surveys (Yan et al. 2008) combined with Census data (Durrant et al. 2008) or questionnaires to the interviewer (Jaeckle et al. 2008); others investigate alternatives to conventional interviewers (Conrad and Schober 2007).

Compared to Germany it seems to be more common for U.S. data collection firms to have interviewers that only work for one particular survey organization (and with that are used to a particular survey house culture), or if they work with others those would be social survey research organizations as well. More importantly it is common for interviewers to be trained centrally from the survey agency at the beginning of their employment and at the beginning of new large scale assignments. Unlike in Germany, face-to-face survey interviewers tend to be paid by the hour no by complete cases. This results in a different incentive structure and also opens the possibility for interviewers to spend time on the aforementioned additional tasks. It goes without saying that the costs of face-to-face surveys are often 10-fold of what is common in Germany.

3. Summary

In summary survey methodologists are conducting new and exciting research into the trade-offs between cost and response rates. As part of these efforts research is done on how best to use non-survey data, either to inform about nonresponse bias or measurement error, but also to supplement data collection and reduce respondent burden. Research is under way to understand error properties of mixed-mode and multiple-frame surveys, but conclusive results are lacking. The German data infra-structure initiative has the possibility to contribute to this research. An overarching theme in all of the above mentioned developments is the increased interest in a relationship between various error sources (Biemer et al. 2008). In Germany there are several good opportunities to engage in research related to the intersection of error sources, especially given the exceptional data linkage efforts that have taken; here Germany is clearly taking the lead compared to the U.S. However, what could be improved in Germany is the collaboration between survey organizations and researchers, the amount of data shared

between those organizations and the willingness to systematically allow for randomized experiments in data collection protocols. In short I would recommend the following:

- Work towards higher quality surveys, especially in the face-to-face field. On step in this direction can be the development of survey methodology standards and the commitment to adhere to these standards. Those standards should include a minimum set of process indicators (metadata), and variables created in the data collection (paradata).
- Expanding options for graduate and post-graduate education in survey methodology could in addition increase the possibility for high quality surveys.
- Carefully examine interviewer hiring, payment and training structures in German survey organizations. Recommendations or minimum requirements regarding these issues might also be needed for German government surveys.
- Use the potential of multiple surveys run within the same survey organizations (or across) for coordinated survey methodology experiments. As we increase the burden on interviewers and try to reduce the burden on respondents many questions will be wide open for the survey methodology research area, such as the effect of question context through matrix sampling, or the effect of interviewer shortcuts when creating sampling frames, or collecting auxiliary data needed for nonresponse adjustment.

References:

Andridge, R. and Little, R. (2008): Proxy Pattern-Mixture Analysis for Survey Nonresponse. University of Michigan. Paper presented at the 2008 Joint Statistical Meeting, Denver, Colorado.

Asef, D. and Riede, T. (2006): Contact times: what is their impact on the measurement of employment by use of a telephone survey? Paper presented at Q2006 European Conference on Quality in Survey Statistics, Cardiff, UK.

Atrostic, B.K./Bates, N./Burt, G. and Silberstein, A. (2001): Nonresponse in US Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. Journal of Official Statistics 17, 209-226.

Axxin et al. (2008): SRC Building New Software to Facilitate Mixed Mode Data Collection Projects. Center Survey, Vol. 18. Issue 6, 8.

Biemer, P./Beerten, R./Japec, L./Karr, A./Mulry, M./Reiter, J. and Tucker, C. (2008): Recent Research in Total Survey Error: A Synopsis of 2008 International Total Survey Error Workshop (ITSEW2008).

Campanelli, P./Sturgis, P. and Purdon, S. (1997): Can you Hear me Knocking: An Investigation into the Impact of Interviewers on Response Rates, London: Social and Community Planning Research (SCPR).

Casas-Cordero, C. (2008): Errors in Observational Data Collected by Survey Interviewers. Comprehensive Examination Paper, Joint Program in Survey Methodology.

Conrad, F. and Schober, M. (2007): Envisioning the Survey Interview of the Future. New York, 298.

Copas, A.J. and Farewell, V.T. (1998): Dealing with non-ignorable non-response by using an enthusiasm-to-respond' variable. Journal of The Royal Statistical Society Series A-Statistics In Society 161, 385-396.

Couper, M. (2005): Technology Trends in Survey Data Collection, Social Science Computer Review, (23), 486-501.

Couper, M. and Lyberg, L. (2005): The use of paradata in survey research. Paper presented at the 54th Session of the International Statistical Institute, Sydney, Australia.

Couper, M./Raghunathan, T./Van Hoewyk, J. and Ziniel, S. (2008): An Application of Matrix Sampling and Multiple Imputation: The Decisions Survey. Paper presented at the 2008 Joint Statistical Meeting, Denver, Colorado.

Curtin, R./Presser, S. and Singer, E. (2000): The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly, (64), 413-428.

De Leeuw, E.D. and De Heer, W. (2002): Trends in household survey nonresponse: A longitudinal and international comparison. In: Groves, R.M./Dillman, D.A./Eltinge, J.L. and Little, R.J.A. (Eds.): Survey nonresponse. New York, 41-54.

De Leeuw, E.D. (2005): To Mix or Not to Mix Data Collection Modes in Surveys. Journal of Official Statistics, Vol.21, No.2, 2005, 233-255.

Dohrmann, S./Han, D. and Mohadjer, L. (2007): Improving coverage of residential address lists in multistage area samples. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Durrant, G.B./Arrigo J. D. and Steele, F. (2008): Using Paradata to Develop Hierarchical Response Propensity Models 19th International Workshop on Household Survey Nonresponse Ljubljana, Slovenia, 15-17 September 2008.

Eckman, S. (2008): Errors in Housing Unit Listing and their Effects on Survey Estimates, Comprehensive Examination Paper, Joint Program in Survey Methodology.

Fricker, S. (2007): The Relationship Between Response Propensity and Data Quality in the Current Population Survey and the American Time Use Survey. Doctoral dissertation. University of Maryland.

Gonzalez, J. and Eltinge, J. (2008): Adaptive Matrix Sampling for the Consumer Expenditure Interview Survey, Paper presented at the Joint Statistical Meeting, American Statistical Association.

Griffin (2008): Measuring Mode Bias in the American Community Survey. Paper presented at the JSPM Design Seminar Spring.

Groves, R.M. (2006): Nonresponse rates and nonresponse bias in household surveys." Public Opinion Quarterly, (70). 646-675.

Groves, R. and Couper, M. (1998): Nonresponse in Household Interview Surveys. 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: Statistics in Society, (169), 439-457.

Groves, R. and McGonagle, K. (2001): A theory-guided interviewer training protocol regarding survey participation. Journal of Official Statistics 17, 249-266.

Groves, R.M., and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly, (72), 167-189.

Groves, B./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. Proceedings of the Survey Research Methods Section, ASA.

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

Groves, R.M./Brick, J.M./Couper, M./Kalsbeek, W./Harris-Kojetin, B./Kreuter, F./Pennell, B./Raghunathan, T./Schouten, B./Smith, T./Tourangeau, R./Bowers, A./Jans, M./Kennedy, C./Levenstein, R./Olson, K./Peytcheva, E./Ziniel, S. and Wagner, J. (2008): Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools. Survey Practice, October 2008 http://surveypractice.org/2008/10/30/issues-facing-the-field/

Heerwegh D./Abts K. and Loosveldt G. (2007): Minimizing survey refusal and noncontact rates: do our efforts pay off? Center for Survey Methodology, Katholieke Universiteit Leuven. Survey Research Methods. Vol. 1, No. 1, 3-10.

Iannacchione, V./Staab, J. and Redden, D. (2003): Evaluating the use of residential mailing addresses in a metropolitan household survey. Public Opinion Quarterly, 67. 202-210.

Jäckle, A./Lynn, P./Nicolaas, G./Sinibaldi, J. and Taylor, R. (2008): Interviewer characteristics, their behaviours, and survey outcomes. Paper presented at the 2008 International Workshop on Household Nonresponse.

Jans, M. (2008): Can Speech Cues and Voice Qualities Predict Item Nonresponse and Inaccuracy in Answers to Income Questions? Prospectus for Dissertation Research, Michigan Program in Survey Methodology.

Keeter, S./Miller, C./Kohut, A./Groves, R. and Presser, S. (2000): Consequences of Reducing Nonresponse in a National Telephone Survey. Public Opinion Quarterly, (64), 125-148.
Kenny-McCulloch, S. (2008): Pretesting Surveys with Traditional and Emerging Statistical Methods: Understanding their Differential Contributions and Affects on Question Performance in Standardized and Adaptive Measurement. Prospectus for Dissertation Research, Joint Program in Survey Methodology.
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). Proceedings of the Survey Research Methods Section, ASA.
Krosnick, J. (2005): Effects of survey data collection mode on response quality: Implications for mixing modes in cross-national studies. Conference on Mixed Mode Methods in Comparative Social Surveys. City University London, 15. Sept. 2005.
Little, R.J.A. and Rubin, D.B. (2002): Statistical Analysis with Missing Data, 2nd Edition. New York.
Lynn, P. (2003): Pedaksi: Methodology for collecting data about survey non-respondents. Quality and Quantity 37, 239-261.
Maynard, D.W./Houtkoop-Steenstra, H./Schaeffer, N.C. and van der Zouwen, J. (2002): Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview. New York.
Mockovak, B. (2008): Mixed-Mode Survey Design. What are the effects on data quality? Challenging Research Issues in Statistics and Survey Methodology at the BLS. U.S. Bureau of Labor Statistics. www.bls.gove/osmr/challening _issues/mixedmode.htm.
Olson, K. (2007): An investigation of the nonresponse – measurement error nexus. Doctoral dissertation. University of Michigan.
Office of Management and Budget (2006): Questions and answers when designing surveys for information collections. www.whitehouse.gov/omb/inforeg/pmc _survey_guidance_2006.pdf.
O'Muircheartaigh, C./Eckman, S. and Weiss, C. (2003): Traditional and enhanced listing for probability sampling. Proceedings of the Section on Survey Research Methods, American Statistical Association.
O'Muircheartaigh, C./English, E./Eckman, S./Upchurch, H./Garcia, E. and Lepkowski, J. (2006): Validating a sampling revolution: Benchmarking address lists against traditional listing. Proceedings of the Section on Survey Research Methods, American Statistical Association.
O'Muircheartaigh, C./English, E. and Eckman, S. (2007): Predicting the relative quality of alternative sampling frames. Proceedings of the Section on Survey Research Methods, American Statistical Association.
O'Muircheartaigh, C. and Heeringa, S. (2008): Sampling Designs for Cross-cultural and Cross National Studies. Paper presented at Multinational, Multiregional, and Multicultural Contexts. Berlin from June 25 - 28, 2008.
Peytchev, A., and Olson, K. (2007): Using interviewer observations to improve nonresponse adjustments: NES 2004. Proceedings of the Survey Research Methods Section, ASA.
Schnell, R. (1997): Nonresponse in Bevölkerungsumfragen. Opladen.
Schnell, R. (2007): Alternative Verfahren zur Stichprobengewinnung für ein IAB-Haushaltspanel mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug. In: Promberger, M. (Ed.): Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“. Nürnberg: IAB Forschungsbericht Nr. 5/2007 .
Schnell, R. and Kreuter, F. (2000): Untersuchungen zur Ursache unterschiedlicher Ergebnisse sehr ähnlicher Viktimisierungs-surveys[KZfSS, 52, 2000. 96-117] Koelner Zeitschrift fuer Soziologie und Sozialpsychologie.
Schnell, R. and Kreuter, F. (2005): Separating Interviewer and Sampling-Point Effects, Journal of Official Statistics, (21), 1–23.
Schouten, B. and Cobben, F. (2007): R-Indexes for the Comparison of Different Fieldwork Strategies and Data Collection Modes. Paper presented at the International Workshop for Household Survey Nonresponse.
Spar, E. (2008): Federal Statistics in the FY 2009 Budget. In: American Association for the Advancement of Science, Report on Research and Development in the Fiscal Year 2009. http://www.aaas.org/spp/rd/rd09main.htm.
Staab, J.M. and Iannacchione, V.G. (2003): Evaluating the Use of Residential Mailing Addresses in a National Household Survey, Joint Statistical Meetings - Section on Survey Research Methods.
Voogt, R.J.J. and Saris, W.E. (2005): Mixed Mode Designs: Finding the Balance Between. Nonresponse Bias and Mode Effects”, Journal of Official Statistics, (21), 367-387.
Wagner, J. (2008): Adaptive Survey Design to Reduce Nonresponse Bias. Doctoral dissertation. University of Michigan.
Yan, T. and Raghunathan, T. (2007): Using Proxy Measures of the Survey Variables in Post-Survey Adjustments in a Transportation Survey, Joint Statistical Meetings - Section on Survey Research Methods.
Yan, T./Rasinski, K./O’Muircheartaigh, C./Kelly, J./Cagney, P./Jessoe, R. and Euler, G. (2008): The Dual Tasks of Interviewers, International Total Survey Error Workshop (ITSEW), North Carolina.