Uwe Engel (Hg.)

Survey Measurements

Techniques, Data Quality and Sources of Error

campus

Survey Measurements

Uwe Engel is Full Professor of Sociology (Statistics and Social Research) at the Department of Social Sciences, University of Bremen, and Head of the Social Science Methods Center at the University of Bremen.

Uwe Engel (ed.)

Survey Measurements

Techniques, Data Quality and Sources of Error

Campus Verlag
Frankfurt/New York

Bibliographic Information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie.
Detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.
ISBN 978-3-593-50280-9

Cover Design: Guido Klütsch, Cologne
Printing office and bookbinder: CPI buchbücher.de, Birkach
Printed on acid free paper.
Printed in Germany

This book is also available as an E-Book.
www.campus.de
www.press.uchicago.edu

Contents

Preface

Uwe Engel

Survey data are error prone and thereby increase the risk of drawing wrong conclusions. Survey methodology therefore pays much attention to possible threats to data quality. Because survey methods are developing further to keep step with a changing world, preserving data quality is a constant task which requires continuing research *on* such methods. To this end, the volume presents recent methodological and statistical research from different countries. The list includes contributions from Germany, Great Britain, the Netherlands, Spain, and the United States. The contributions focus on different sources of survey error and present techniques to cope with their negative effects on survey measurements.

The present volume relates to the Priority Programme on Survey Methodology (PPSM) of the German Research Foundation (DFG). On the one hand, it reports on research which has been conducted as part of this Programme. On the other hand, it reports on research which was carried out by survey methodologists and statisticians who accompanied this Programme scientifically; they include leading researchers who held invited speeches at PPSM conferences and helped to identify future directions in survey methodology at an international workshop.

The Priority Programme on Survey Methodology commenced in January 2008. Sixteen projects were undertaken over the course of the following six years. The coordination project organized three biennial conferences for an international audience in 2009, 2011, 2013, and an international workshop in 2014. A related volume to the present one is published elsewhere to report on relevant scientific work as extensively as possible (Engel et al. 2015).[1]

For whom is this volume written? We hope that the volume will be of benefit to three primary audiences. First of all, it will assist applied survey

1 Engel, U., Jann, B., Lynn, P., Scherpenzeel, A., and P. Sturgis (eds.) (2015). *Improving Survey Methods: Lessons from Recent Research.* New York: Routledge.

researchers in designing their survey studies at a state-of-the-art level. Survey statisticians and survey methodologists, in their roles as both researchers and teachers, represent another audience. The book should also be appropriate as course reading at the advanced B.A., M.A., and Ph.D. level in university departments that offer specialized courses on survey methods to their students.

I am grateful to all authors for their excellent contributions. Special thanks go to Laura Burmeister, Sabine Sommer, and Jennifer Wessels for their always excellent organizational assistance and to Katherine Bird who checked all chapters linguistically from the point of view of an English native speaker, to make final linguistic amendments when necessary.

Regarding the Priority Programme on Survey Methodology (PPSM), the financial support this programme received from the German Research Foundation (DFG) is gratefully acknowledged. Regarding the PPSM panel, special thanks go to the University of Bremen as well.

Bremen, March 2015
Uwe Engel

1. Introduction

Uwe Engel

1.1 Data Quality

Surveys are important for society. They are frequently conducted and useful sources of public opinion and decision making. Although even outcomes of high-quality surveys are not safe from being misinterpreted, either inadvertently or even deliberately, high-quality survey data are likely to reduce this risk. For scientific reasons as well, strictly speaking only high-quality survey data appear acceptable. This is why survey methodology pays so much attention to possible threats to data quality and has been doing so for quite some time (e.g. Biemer and Lyberg 2003, Weisberg 2005).

Why is high data quality so important for survey research? One possible answer to this question may point to the risk of obtaining biased sample estimates of population parameters if a survey fails to cope with relevant sources of survey error. Probability sampling and proper use of statistical estimators alone cannot guarantee unbiased estimates, because even in this case nonresponse and measurement effects may still give rise to bias and error variance.

Accordingly, one core task certainly consists in the development of suitable statistical models and techniques to adjust for nonresponse bias. Even the ideal case of complete (or perfectly nonresponse-adjusted for) response, however, cannot guarantee unbiased samples estimates for a simple reason: Observed responses may deviate from their corresponding true scores due to measurement effects.

Such effects may have different origins, including the survey mode, question wordings, and response formats. In addition to such 'mode' and 'response effects', the 'interviewer' represents a further source of measurement error. Of importance is also the 'respondent' insofar as his/her response behavior may differ in relevant aspects. In this respect, typical examples are certainly satisficing behavior and cognitive response styles. Another source of variation is simply that respondents arrive at their responses to survey questions through cognitive processes that may differ in relevant regards. No less important than this, however, is another factor of answering behavior which might be called 'motivated misreporting'.

A working definition of 'high-quality' surveys might thus include the idea that the quality is the higher the more such sources of survey error are effectively controlled for. In doing this, one would adopt the prominent 'total survey error' perspective.

1.2 Sources of Survey Error

1.2.1 Measurement Error

Measurement error may be due to several sources of variation that affect response behavior. Surveys do *not* yield unobtrusive measurements. Instead, already the fact *per se* that respondents are asked questions in the context of research interviews shapes their answering behavior in some ways.

Survey-mode effects

It is well known that different survey modes produce different mean values, other things being equal. It makes a difference whether a finding has been obtained in an interviewer-assisted or self-administered survey mode. For instance, the analysis presented in chapter 10 below exemplifies the typical observation that the web mode tends to produce lower mean values than the telephone mode. 'Lower' means at the same time 'farther away' from an answer the respondent is assumed to believe to be an expected, i.e. socially desirable, answer. In the aforesaid analysis, this is the assumed expectation of presenting oneself as currently satisfied with one's life. If posed in direct communication, the mere presence of an interviewer gives rise to a kind of 'positivity bias' (Tourangeau et al. 2000, 240f.) and this bias in turn to a comparatively higher mean value than observed in the opposite case of self-administered survey modes. This is just an example of a kind of measurement effect which is usually termed 'mode effect'.

Response effects

Other measurement effects are called response effects and evolve from the way questions are worded and response formats are styled. Experimental research shows, for example, that different response distributions arise from different response formats of closed-ended survey questions, other things being equal (e.g. Engel et al. 2012, 286ff.). In the present volume, particular attention is paid to open-ended questions and possible framing effects.

Open-ended questions

Open-ended questions allow the formulation of answers in the respondent's own words. This leads to more or less content and thus to the need of properly analyzing this content. Nowadays, *content analysis* certainly ranks as one of the methods of growing importance in social research. Not only the sheer amount of content provided through web sites and social media is likely to contribute to this development. The analysis of open-ended questions in surveys is a challenging task, too. This becomes evident from the fact that verbatim answers represent more or less unstructured text material from which the survey researcher has to extract meaningful information and structure. In this respect, the usual approach is *theory-driven* and implies having to master the task of coding the answers properly. Accordingly, there exists a strong research interest in accomplishing this task as error-free as possible. For this reason, additional insights into the structure of verbatim answers may be gained by complementing this theory-driven approach to coding verbatim responses by *data-driven* techniques of revealing hidden structures.[1] In the present volume, however, the challenge preceding any coding attempt is not addressed.

Chapter 3 deals with open-ended survey questions. First of all, *Sturgis* and *Luff* discuss some merits of this type of question (e.g. allowing the respondent to use his or her own frame of reference in answering a question and the potentially rich informational value of answers to open-ended questions). The authors discuss the role of interviewers as potential sources of error, because interviewers "must type the verbatim answer as the re-

1 We think of automated text mining methods. In particular, we think of machine-augmented analysis of textual material (e.g. Haney 2014; Keyling 2014), which includes the use of reference textual corpora and via this route a dynamically increasing body of known and thus meaningfully analyzable strings (machine learning).

spondent articulates it, often in less than ideal conditions." This makes interviewer transcription, which is the central chapter topic, error prone. The chapter therefore discusses an alternative to letting the interviewers type in verbatim responses. This is 'audio-recording' the responses to open-ended questions (OEQs). As the authors note, "in this chapter we assess the costs and benefits of audio-recording responses to OEQs in the context of a computer-assisted personal (CAPI) survey." Based on random allocations of respondents to the conditions 'audio-recording' versus 'interviewer-typed' in the *2012 Wellcome Trust Monitor* survey, the authors examine the data quality in both conditions and discuss audio-recording also with respect to the necessary consent to be audio-recorded.

Open- and closed-ended survey questions combined

From its beginnings, social research has combined different methods. Nowadays, we observe a growing recognition of the idea of 'mixing' methods. Other than the 'mixed-mode' parlance which is so popular in current survey methodology, the talk is usually of 'mixed methods' in order to designate efforts of combining specifically 'qualitative' and 'quantitative' methods. Applied to the narrower survey methodology field and there to *within-survey* applications, open-ended questions in usually standardized surveys may be regarded as a potential field of application. In this respect, the combination of closed-ended survey questions with relevant open-ended *meaning probes* and *think-aloud probes* may prove particularly promising. 'Probing' is by no means a new questioning technique, quite the contrary. It is only remarkable that its 'traditional' place is the pretesting stage of surveys. However, 'probes' are simply *meta*-questions (in the sense of questions *about* questions) which we can pose theoretically in the current surveys as well. They are meta-questions pertaining to given questions, in order to clarify how respondents interpret these questions and how they arrive at their answers to these questions. We explored the feasibility of this approach elsewhere (Engel and Köster 2015, 45–47) and were led to find it promising.

Framing effects

Surveys very often employ question formats that fully standardize both question wordings and answering formats. The comparability of answers is certainly the primary merit of such a full standardization. There are, however, two sides to the coin. If 'comparability' represents one such side, 'framing' represents the other. Comparable answers can always be obtained only relatively to the frame implicitly set by a question. Since, with a closed-ended question, people can only give one of the answers that they have been offered, we would obtain other answers if we'd change the framing with respect to these response categories or *if we do without any such category*.

The replacement of closed-ended with open-ended answering formats may indeed reveal a framing effect. *Chapter 4* reports on an experiment that yields evidence of such a *response frame* effect. Furthermore, *Engel* and *Köster* show that framing effects may also be caused by *lead texts* of survey questions. Such effects become visible in framing experiments in surveys in which random halves of samples are asked questions about the same topic, "but following lead-ins that frame the issue in different ways" (Weisberg 2005, 121). Chapter 4 uses a specific factorial survey design to conduct one such experiment. 'Specific' means that the comparison is embedded in a real (i.e. not only hypothetical) experimental structure.

Motivated misreporting

Measurement effects may be due to factors like survey modes, question wordings, response formats, question and response orders. Explanations of such effects may also require simultaneous consideration of respondents' characteristics. Mode effects, for example, may evolve from the tendency to respond to survey questions in a way which the interviewee believes the interviewer expects to hear. 'Social desirability' is sometimes used to designate the pole toward which survey responses may be biased. It appears thus meaningful to assume that the involvement vs. non-involvement of an interviewer is a relevant factor. The aforesaid motivational tendency, however, is likely to vary also as a function of individual characteristics, for example personality traits.

The 'sensitivity' of survey questions is another case in point. To minimize the probability of provoking biased answers in face-to-face interviews, for example, it is common survey practice to switch to a self-administered mode for asking sensitive questions. Another example of coping with sensitive questions is the employment of special questioning techniques (Jann 2015). However, when is a survey question sensitive and when it is not? Because it is believed that sensitive questions may give rise to item nonresponse, the latter may be taken as an indicator of sensitivity. 'Income' is a typical example. However, experience teaches that quite different information may be regarded as sensitive if the item-response criterion is taken as benchmark. Age, household structure, and family status, for example, and thus information which might appear pretty harmless at first glance (Engel 2013, 68f., Tables 3.23 and 3.24). The story is simply that people may find quite different information 'too personal' to be asked in a research interview. If the degree of being *too personal* is asked explicitly via relevant meta-questions, then an interesting relationship becomes evident. Namely that this *perceived* sensitivity may systematically depend on the behavior in question. We observed an instance of this relationship in an analysis of response effects described elsewhere (Engel and Köster 2015, 41). For, among other things, this analysis revealed that the perceived sensitivity of a question about the frequency of alcohol consumption was a monotone function of exactly this frequency. This finding *per se* is probably not so surprising. However, it let us assume that the answers *which are reported in survey interviews* do not necessarily reflect true scores but those scores which appeared just acceptable to the interviewees. What we get reported, were accordingly *motivated* answers.

A decisive question is accordingly whether respondents tend to report true values or whether they prefer to deliberately report just acceptable deviations from such true values. Motivated misreporting is certainly a relevant topic.

Chapter 2 addresses the topic of 'motivated misreporting' in a clearly specified sense. As *Tourangeau*, *Kreuter,* and *Eckman* outline in their introduction, one reason for misreporting is "that the respondent wants to avoid making potentially embarrassing revelations to an interviewer". They point to two typical examples, namely that "survey respondents underreport their use of illicit drugs and overreport having voted". The authors continue making clear that their chapter examines "another reason why respondents may give distorted answers – the respondents are motivated to misreport because they want to shorten the interview and avoid additional burden." The chapter examines the evidence that motivated misreporting may contribute to three forms of measurement error in surveys (referring to screening questions, filter questions, and panel conditioning).

1.2.2 Nonresponse Error

It is common to distinguish between two basic forms of nonresponse, unit nonresponse, and item nonresponse. 'Unit nonresponse' stands for the loss of whole 'units', i.e. for the loss of information of complete target persons of a survey. There are always target persons who surveys fail to reach or, given contact, fail to convince to participate. Unit nonresponse is challenging because of the inherent risk of leading to sample compositions that deviate systematically from the compositions of the parent populations the random samples have been drawn from. This is why 'unit nonresponse' is likely to cause bias in sample estimates and this is why appropriate adjustments for unit nonresponse may really be necessary.

In panel surveys, unit nonresponse is a phenomenon that is likely to occur not only on the first measurement occasion but also at downstream panel waves. There are always respondents who drop out of a panel study after having participated at least once. Experience teaches that despite all efforts attempts at re-contacting and re-interviewing panel members are not always successful, that way giving rise to so-called 'panel attrition'. Panel-based estimates of change are then confounded unless true change is separated from systematic dropout effects. An example is presented in *chapter 10*.

Unit nonresponse may also be due to another factor than failure to achieve contact and cooperation with target persons. Surveys over the internet, for example, can only reach people with access to the web. Surveys using mobile devices like smartphones and tablet computers cannot reach people who do not have such devices at their disposal. Unit nonresponse may therefore result from non-coverage. In this connection a related factor is the readiness to use such devices if requested. It is easily imaginable that people *do have* access to a relevant mobile device, say a smartphone, but are not prepared to use it for answering survey questions via this particular tool. Survey methodology thus tries to understand the factors underlying this variant of survey cooperation. Referring to the willingness to take part in an experience sampling study using mobile devices, *chapter 9* analyzes such factors more closely. There, *Burmeister*, *Engel,* and *Schmidt* describe the experience sampling method and report findings about the willingness to participate in such a study.

'Item nonresponse' designates the second basic form of nonresponse. There are always people who take part in requested interviews but refuse to answer every survey question. This leads to incomplete data matrices that the survey researcher has to cope with. An overview of relevant missing data techniques is given in *chapter 11*.

1.2.3 Web Panels and Mobile Web Surveys

Preserving data quality in survey research is a constant task. Survey methods are developing further to keep step with a changing world. Information technology creates new modes of communicating with each other. Where these modes are adopted on a larger scale in society, survey research has to react to such changing habits of interpersonal communication for two basic reasons: to maintain access or acquire better future access to populations of interest. Accordingly, technical progress and changing communication habits give rise to emerging new survey methods and shifts in the relative importance of established ones. Nowadays, telephone and face-to-face surveys compete with web panels, while in the future the mobile web is likely to gain in importance (Callegaro et al. 2014). Not least, these trends favor the mixture of different survey modes, either sequentially or concurrently, and create new challenges survey methodology has to cope with. The identification and proper handling of threats to data quality consequently requires continued research *on* survey methods.

Chapter 7 deals with quality issues of web panels. First of all, *Bethlehem* discusses weaknesses of web *surveys* in view of methodological issues related to the probability sampling paradigm (undercoverage, sample selection, nonresponse). Mixed-mode surveys are considered one possible solution, web *panels* another. It is pointed out that such panels may be used for doing longitudinal research or as a sampling frame for specific surveys, i.e. as an access panel. The principal point then is that web panels too, and even probability-based ones, may suffer from selection problems. The chapter explores these problems in detail and offers solutions of how to correct selection bias.

The chapter reports on the *Web Panel Pilot of Statistics Netherlands* and concludes the discussion with specifying the conditions under which a web panel can be a useful tool. It is this context in which the chapter also addresses the 'measurement error' topic in some detail.

Chapter 8 addresses the topic of mobile responding and, as the authors note, should "be viewed in the context of existing web panels and regular web surveys which have not been especially adapted for mobile devices." *De Bruijne* and *Oudejans* refer to the situation that surveys designed for computers are also being completed using mobile browsers. The authors consequently raise the question "of how mobile response possibly affects the survey results and data quality." The chapter presents a theoretical background of the cognitive response process. Based on a recent internet survey in the Dutch *CentERpanel*, the authors explore in particular the situational context of mobile responding and conclude from their findings that online respondents "are no longer confined to a computer-based, quiet, high-focus environment. Online surveys have entered the living room and are taken by many amidst daily life." The authors continue in highlighting the impact of the situational context on the respondent's level of attention and cognitive processing when answering survey questions.

1.3 Techniques

1.3.1 Combining Survey Methods

It is expected that mixed-mode designs will become increasingly common in future survey research as a means of counteracting declining response rates that differ across different segments of the population (de Leeuw and Hox 2015; Massey and Tourangeau 2013; Kreuter 2013; Stoop 2015). For example, with mixed-mode designs, researchers tailor the contact process, target the use of incentives, and give respondents a (sequential) choice of data collection mode. However, given the multiplicity of survey modes employed in survey practice, it will become increasingly important to identify and adjust for possible 'mode effects' (de Leeuw, Hox, and Dillman 2008, de Leeuw and Hox 2011). This will become all the more important as more combinations of survey modes and devices are emerging in research.

In addition, the combination of methods may follow the 'mixed-method' perspective in integrating more closely qualitative and quantitative types of analyzing survey data.

1.3.2 Coping with Nonresponse Error

Missing data represent a ubiquitous phenomenon in survey research. Since unit nonresponse is likely to cause bias in sample estimates, appropriate adjustments for nonresponse are highly advisable. From a statistical point of view, this task may be realized by computing, for example, propensity scores using probit or logistic regression models which may be used for the computation of nonresponse weights.

From a methodological point of view, the task is quite complicated. Response propensities are estimated response probabilities whose estimation requires information about both respondents *and* nonrespondents. Address-based sampling and the accompanying access to related information does certainly help accomplishing the task of obtaining information about persons who prove nonrespondents later on. However, even the use of address-related register and area information is usually of only limited suitability in attempts at adjusting for bias *in survey variables* (because the latter variables may be correlated only weakly with the external register and area data). It thus appears advisable to replenish such data with additional auxiliary information when estimating individual response probabilities. In this respect, relevant developments are discussed using umbrella terms like 'paradata' (Kreuter 2013a; Kreuter and Olson 2013) and 'adaptive' survey designs (Bethlehem et al. 2011; Engel 2015).

Several statistical missing data techniques exist. *Chapter 11* provides an overview of such techniques and possible missing data patterns. *Salfrán* and *Spiess* discuss the so-important 'ignorability' of the missing mechanism. The chapter deals with ad-hoc methods, ML estimation, weighting, and multiple imputation techniques.

Chapter 12, too, deals with 'multiple imputation' as a statistical technique to handle missing data. Today multiple imputation is certainly among the standard methods of dealing with missing data. Despite its popularity and the implementation of relevant procedures in most data analysis packages, *Kleinke* and *Reinecke* have to point out that "currently available commercial statistical software is still highly limited regarding the imputation of incomplete count data, and especially multilevel count data." The authors thus fill a gap by having developed a solution "to create multiple imputations of incomplete overdispersed multilevel count data". The procedure is based on the "multiple imputation by chained equations approach" and works as an add-on to the mice software in R. The chapter gives a detailed exposition of this procedure and its foundations.

1.3.3 Coping with Measurement Error

Statistical approaches

Nowadays, measurement error is handled routinely. The structural equation and mixture modeling framework is almost designed for such a purpose (Muthén and Muthén 1998–2010). For *continuous* latent variables, for example confirmatory factor analysis (CFA) may be used as a standard tool for the specification and estimation of models that consider random and systematic measurement error (Byrne 2012).

While the consideration of random measurement error represents the standard approach, models are easily extended to include systematic measurement error as well. Relating to the latter, typical cases include the repeated exposure of respondents to one and the same measuring instrument, for instance when in panel studies respondents are requested to answer the same survey questions several times. This practice gives rise to a special kind of 'method effect'.

This effect may be considered implicitly by setting free relevant residual correlations (as for instance is done in chapter 10), while an alternative strategy consists in estimating these and other methods effects explicitly via so called MTMM models. 'MTMM' is the usual acronym for 'MultiTrait-MultiMethod' and may, nowadays, be conceived of as the MTMM approach *to confirmatory factor analysis*.

This special CFA approach helps realizing the core idea of achieving estimates of reliability and validity which are freed of possible confounding method effects. A typical case arises if each of a set of relevant 'traits' (latent variables) is measured alike on the basis of different response formats, i.e. on the basis of 'multiple methods'.

Chapter 5 shows how to use a split-ballot MTMM approach to assess the quality of differently labeled frequency scales. *Revilla* and *Saris* examine in particular the hypothesis that the choice of exact scale labels has an impact on the quality of these scales. The authors introduce their approach while reporting on findings of MTMM experiments using recent data from the Online Panel *Netquest* for Spain, Colombia, and Mexico.

Chapter 6, too, deals with the MTMM approach. Here, the different 'methods' are 2pt, 4pt, and 11pt scales. *Burmeister* and *Engel* address a specific question which may come up in fieldwork for MTMM research. MTMM modeling may require respondents to answer pretty similar questions just to obtain the different 'methods' needed for an MTMM analysis. Since respondents may find such repeated questions confusing and even annoying, the suitability of a special questioning approach to the collection of MTMM data was examined, namely a branching format without filtering.

For *categorical* latent variables, Latent Class Analysis (LCA) represents a powerful statistical tool to control for measurement error. Latent class analysis of survey error is described elsewhere in greater detail (Biemer 2011; see also Alwin 2007, 263ff.). In the present volume, *chapter 10* illustrates LCA modeling using the example of a latent transition analysis (LTA). Specifically a Latent Markov Model for two measurement occasions is estimated and then refined in two basic ways to consider the Mover-Stayer distinction and the effect of panel dropout.

Statistical and cognitive approaches to error combined

In their seminal work published one and a half decades ago, Tourangeau et al. (2000, 321f.) suggested[2] "the possibility of combining statistical models' methods of partitioning error with the cognitive models' ideas about the sources of these errors." Such a combination may take quite different forms. For example, interviewer ratings of satisficing behavior may be employed to assess the impact of satisficing *on sample estimates* which are adjusted for other forms of measurement error and nonresponse error at the same time. A latent growth curve approach was pursued to consider the *direct* effects of satisficing, response propensity, survey mode, mode preference, and answering scale format (Engel 2013, 92–96). In another context, regression analysis of experimental data on response order (frequency scale with categories sorted in ascending vs. descending order) considered the conditional effects of satisficing behavior and perceived sensitivity on the strength of a response-order effect in question (Engel and Köster 2015). The relationship between three cognitive response styles, as identifiable via Latent Class Analysis (answers anchored in end-

2 With reference to Groves.

points with/without an affinity to grade one's opinion, answers anchored in the middle category of a scale), and reluctance and satisficing behavior represents a further relevant example (Engel and Köster 2015).

In addition to direct measurement effects, *indirect* effects matter, too. The core intermediary element is the individual 'response propensity'. Such indirect measurement effects may be considered if a survey design enables the researcher to estimate response propensities *longitudinally*. This applies, for example, to probability-based access panels because of the possibility of computing propensity scores at each of a series of consecutive selection steps. In doing this, except for the first such step (preceding the recruitment interview itself), the prediction equation of downstream response probabilities may include relevant ratings of the response behavior and the interview situation.

Chapter 10, for instance, illustrates this possibility in the shape of one of the prediction equations involved in the latent growth curve analysis presented there. Specifically, the probability of *expressing readiness* to join the panel proved to be influenced by the satisficing degree of response behavior, the sensitivity of survey items, and the perceived interview atmosphere (Engel 2013, 55f., Table 3.3). Since a propensity score represents the bundled impact of all predictor variables that went into its computation, the propensity score in question thus in part reflects the impact of aforesaid variables too when, in turn, the impact of this propensity score on survey estimates is estimated.

This is what *Can* and *Engel* carry out in the analysis reported in chapter 10. They start with a Curves of Factor Model to estimate an initial value and the expected linear change over time net of measurement error. They then expand the model to a Pattern-Mixture model, to consider in addition dropout effects on the estimate of change. Finally, the model is enlarged again by also considering differential response propensities evolving from factors that precede the actual panel.

References

Alwin, D. F. (2007). *Margins of Error. A Study of Reliability in Survey Measurement.* Hoboken NJ: Wiley.

Bethlehem, J., Cobben, F., and B. Schouten (2011). *Handbook of Nonresponse in Household Surveys*. Hoboken: Wiley.

Biemer, P. P. and L. E. Lyberg (2003). *Introduction to Survey Quality*. Hoboken, NJ: Wiley.

Biemer, P. P. (2011). Latent Class Analysis of Survey Error. Hoboken, NJ: Wiley.

Byrne, B. M. (2012). *Structural Equation Modeling With Mplus.* New York: Routledge.

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., and P. J. Lavrakas (eds.) (2014). *Online Panel Research. A Data Quality Perspective*. Chichester: Wiley.

De Leeuw, E. D., Hox, J. J., and D. A. Dillman (2008). Mixed-mode Surveys: When and Why. In E. D. de Leeuw, J. J. Hox, and D.A. Dillman (eds.). *International Handbook of Survey Methodology,* 299–317. New York: Lawrence Erlbaum Associates.

De Leeuw, E. D. and J. J. Hox (2011). Internet Surveys as Part of a Mixed-Mode Design. In M. Das, P. Ester, and L. Kaczmirek (eds.*). Social and Behavioral Research and the Internet. Advances in Applied Methods and Research Strategies*, 45–76. New York: Routledge.

De Leeuw, E. D. and J. J. Hox (2015). Survey Mode and Mode Effects. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis (eds.). *Improving Survey Methods: Lessons from Recent Research*, 22–34. New York: Routledge.

Engel, U. (2013). *Access panel and mixed-mode internet survey. PPSM Panel Report*. 17.02.2015 http://www.sozialforschung.uni-bremen.de/html/downloads.html.

Engel, U. (2015). Response Behavior in an Adaptive Survey Design for the Setting-Up Stage of a Probability-Based Access Panel in Germany. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis (eds.*). Improving Survey Methods: Lessons from Recent Research*, 207–222. New York: Routledge.

Engel, U. and Köster, B. (2015). Response effects and cognitive involvement in answering survey questions. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis (eds.) *Improving Survey Methods: Lessons from Recent Research*, 35–50. London: Routledge.

Engel, U., Bartsch, S., Schnabel, C., and H. Vehre (2012). *Wissenschaftliche Umfragen. Methoden und Fehlerquellen*. Frankfurt: Campus Verlag.

Haney, C. (2014). Sentiment Analysis: Providing Categorical Insights into Unstructured Textual Data. In C. A. Hill, E. Deana, and J. Murphy (eds.). *Social Media, Sociality, and Survey Research.* Kindle eBook Edition, Chapter 2 (Position 1163–1636). Hoboken: Wiley.

Jann, B. (2015). Asking Sensitive Questions: Overview and Introduction. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis (eds.). *Improving Survey Methods: Lessons from Recent Research*, 101–105. New York: Routledge.

Keyling, T. (2014). Automatisierte Inhaltanalyse. In M. Welker, M. Taddiken, J. -H. Schmidt, and N. Jackob (eds.). *Handbuch Online Forschung. Sozialwissenschaftliche Datengewinnung und –auswertung in digitalen Netzen*. Köln: Herbert von Halem Verlag.

Kreuter, F. (2013). Facing the Nonresponse Challenge. In S. Massey, and R. Tourangeau (eds.). *The ANNALS of the American Academy of Political and Social Science*, 645, 23–35.

Kreuter, F. (2013a). Improving Surveys with Paradata: Introduction. In F. Kreuter (ed.). *Improving Surveys with Paradata. Analytic Uses of Process Information,* 1–9. Hoboken: Wiley.

Kreuter, F. and K. Olson (2013). Paradata for Nonresponse Error Investigation. In F. Kreuter (ed.). *Improving Surveys with Paradata. Analytic Uses of Process Information,* 13–42. Hoboken: Wiley.

Massey, D. S. and R. Tourangeau (2013). Where Do We Go from Here? Nonresponse and Social Measurement. In D. S. Massey and R. Tourangeau (eds.). *The ANNALS of the American Academy of Political and Social Science* Vol. 645, 222–236.

Muthén, L. K. and Muthén, B. O. (1998–2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén and Muthén.

Stoop, I. (2015). Nonresponse in Comparative Studies. Enhancing Response Rates. Detecting and Minimizing Nonresponse Bias. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis (eds.). *Improving Survey Methods: Lessons from Recent Research,* 351–362. New York: Routledge.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Weisberg, H. F. (2005). *The Total Survey Error Approach. A Guide to the New Science of Survey Research*. Chicago/London: The University of Chicago Press.

2. Motivated Misreporting: Shaping Answers to Reduce Survey Burden

Roger Tourangeau, Frauke Kreuter, and Stephanie Eckman[1]

2.1 Introduction

In their model of the survey response process, Tourangeau, Rips, and Rasinski (2000) noted that respondents may 'edit' their answers before they report them. One reason for such editing is that the respondent wants to avoid making potentially embarrassing revelations to an interviewer. For example, survey respondents underreport their use of illicit drugs and overreport having voted (Tourangeau and Yan 2007). In this chapter, we examine another reason why respondents may give distorted answers – the respondents are motivated to misreport because they want to shorten the interview and avoid additional burden. Furthermore, interviewers might also be motivated to shorten the interview, and thus elicit (or record) inaccurate answers. We examine the evidence that motivated misreporting may contribute to three forms of measurement error in surveys.

The first type of measurement error that may be a product of motivated misreporting occurs with questions designed to identify members of the eligible population for a survey (or 'screening' questions, in survey parlance). There is considerable evidence that screening interviews generally miss some members of the target population (see, for example, Horrigan et al. 1999 and Judkins et al. 1999). Of course, virtually all surveys miss *some* members of their target populations, even if there is no screening for eligibility but simply a rostering of all persons living in the household. For example, Fay (1989) provides estimates of the within- household under-

1 Much of the work discussed here was supported by the National Science Foundation (NSF) under Grants SES 0850999 (to Frauke Kreuter) and 0850445 (to Roger Tourangeau). The authors would like to thank the Methodology, Measurement, and Statistics Program of the NSF and Dr. Cheryl Eavey for their support. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.