Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
SOC2012-0029

Measuring Nonsampling Errors: A Challenge to the Total Survey Error Future Research Agenda

Darina Peycheva
PhD Student
Social Survey Research Center,
Institute for the Study of Societies and Knowledge,
Bulgarian Academy of Sciences
Bulgaria

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209
Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

Printed in Athens, Greece by the Athens Institute for Education and Research.
All rights reserved. Reproduction is allowed for non-commercial purposes if the source is fully acknowledged.

ISSN **2241-2891**

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Peycheva, Darina (2012) "**Measuring Nonsampling Errors: A Challenge to the Total Survey Error Future Research Agenda**" Athens: ATINER'S Conference Paper Series, No: SOC2012-0029.

Measuring Nonsampling Errors: A Challenge to the Total Survey Error Future Research Agenda

Darina Peycheva
PhD Student
Social Survey Research Center, Institute for the Study of Societies and Knowledge, Bulgarian Academy of Sciences
Bulgaria

Abstract

The paper outlines the role of the `*Total Survey Error*` (TSE) paradigm in the framework of the `*Survey Quality*` concept, while focusing on its key weaknesses in terms of measuring major TSE components in practice. Taking into account that the measurement of the TSE components and especially the measurement of nonsampling errors is still a challenge to the survey researchers and a main priority of the TSE future research agenda, the paper discusses the methods measuring specific nonsampling errors and evaluates their limitations in terms of quantifying the magnitude of the sources of error on the survey output quality.

The background of this paper is a study on the evolution of the `*Survey Quality*` concept that shows the TSE is the conceptual foundation of the field of survey methodology and the core of each survey quality perspective. However, none of the perspectives towards survey quality have overcome the limitations of the difficult (quantitative) measurement of some of the TSE components. Thus the measurement of the TSE components is still a challenge to the survey methodologists. Nevertheless, the decomposition of errors done in the article allows studying each TSE component in details, thus extends the knowledge on their effect and makes possible the development of measures to limit their magnitude on the survey data quality.

Keywords: Total survey error (TSE), Survey Quality, Data Quality, Nonsampling Errors, Nonsampling Errors Measurement

Acknowledgements: The author thanks to Corr. Mem. Prof. Atanas Atanasov (BAS) and Dr. Ekaterina Markova (Social Survey Research Center, ISKS-BAS) for the valuable comments and suggestions, and the stimulating discussions in Survey Methodology prodding in getting the paper done.

Contact Information of Corresponding author: D. N. Peycheva, Social Survey Research Center, Institute for the Study of Societies and Knowledge, Bulgarian Academy of Sciences, 13A Moskovska St., 1000 Sofia, Bulgaria (e-mail: dpeicheva@gmail.com).

Intorduction

A comprehensive study on the evolution of the `survey quality` concept shows that the total survey error (TSE) paradigm is the core of each survey quality perspective. Moreover `survey quality` appears as an ongoing improvement concept. Each following quality perspective continuously embraces new dimensions deriving from the limitations of the previous ones. However, in the attempt to embrace more and new quality indicators, the different survey quality perspectives have not overcome the weakness of the core quality indicator set up in the TSE framework – the accuracy of the survey outputs. For many, accuracy means the measurement of sampling error but, in fact, the concept is much broader, taking into account the nonsampling error as well. Nonsampling error includes coverage error, measurement error, nonresponse error, and processing error. Some of these nonsampling errors are still a challenge to survey researchers and methodologists in terms of measuring their magnitude on the survey output. Nevertheless, even the attempt to decompose the nonsampling error into its specific components places the evaluation process a step forward as these become easier to be studied and managed.

Here we provide an overview on the decomposition of the nonsampling error into more manageable error components and discuss approaches towards their measurement. We outline the extent these approaches give a quantitative evaluation on the magnitude of the specific error on the survey output.

1. **TSE paradigm in the framework of the `Survey quality` concept: key strengths and weaknesses**

'*Survey quality'* concept can be described from different perspectives (e.g. `*Fitness for Use*`, `*Quality at Three Levels*`, `*Total Quality Management*`, etc.). The selection and application of one or another quality perspective is always an organizational (management) choice. No matter which quality approach is chosen, it is always important that the organization strives to improve the quality and to better manage its surveys.

Although various survey quality perspectives exist, the study on their evolution shows that each following perspective steps on the foundation of the previous one, while trying to overcome its limitations and to go beyond. However, the `*total survey error*` (TSE) paradigm is the core of each survey quality perspective and it is the conceptual foundation of the field of survey methodology. Thus the evolution of the `survey quality` concept tightly follows the chronological and thematic development of the TSE paradigm (Groves & Lyberg, 2010). This conceptual framework describes the statistical characteristics of survey estimations while integrating the different sources of errors, and it does not include any non-statistical indicators set up frequently in the assessment of survey data by other quality perspectives. The explicit attention to survey errors is the undeniable strength of this concept. However, the TSE effect on data accuracy is hardly (quantitatively) measured. Some of the TSE components, such as the margin of sampling error, are relatively easily calculated and familiar to many who use survey research. Other components, such as the undercoverage effect, the characteristics of nonrespondents, the influence of question wording on response, or interviewer effect, are more difficult to ascertain (AAPOR, 2011). The impossibility of quantitative measurement of most of these components in practice is the key weakness of the framework (Biemer, 2010). Nevertheless, the decomposition of errors allows studying each TSE component in detail and thus extends the knowledge on their effects and makes possible the development of measures to limit their magnitude on the survey output quality. As this limitation of the TSE framework has not been overcome by any other quality perspectives, the nonsampling error measurement is still the challenge to the future survey research agenda.

2. Measuring nonsampling errors: capabilities and limitations of the methods and tools

It is clearly said in the methodological literature that in order to achieve a survey with high quality we need to reduce the TSE[1] to the most possible extent. However, the idea is not to generally reduce the TSE, but to decompose it into specific error components and then to further subdivide it into smaller and more manageable sources (Beimer, 2010). This makes the process of studying each survey error source easier and more detailed, while simultaneously improving the approaches towards their measurement. However, the TSE concept itself is not as clearly defined as it seems, since the different methodologists include different components of error in it for example. In survey methodology we can meet different typologies and classifications of errors. *Groves (1989)* presents a detailed classification of survey errors in the framework of the TSE, where the nonsapling errors are catalogued into three potential areas in which error can occur in sample survey. One is the coverage, where error can result if some members of the population under study do not have a nonzero chance of being included in the sample. Another is the measurement effect, such as when the instrument or item on the instrument is constructed in such a way as to produce unreliable or invalid data. The third is the nonresponse effect, where nonrespondents in the sample that researchers originally draw differ from respondents in ways that are relevant to objectives of the survey (AAPOR, 2011). However, in his classification Groves does not include processing errors.

Here we concentrate our attention on the most common approach to decompose nonsapling errors into coverage error, nonresponse error, measurement error, and processing error (Czaja & Blair, 2005; EHQR, ESQR, 2009; US Statistical Policy, 2001). Different aspects related to these types of error are of significant importance to survey methodologists, for example how do they arise, what are the methods and procedures to minimize their effects, etc. But what is considered to be important in the frame of the current paper are the methods aiming to measure their magnitude on the survey output. Sometime this measurement has quantitative results (e.g. rates, indexes, etc.), but more often the magnitude of specific error components cannot be directly measured or expressed with quantitative estimates.

2.1 Measuring Coverage Error

Coverage error (also known as frame error) is the difference between the population and the sample frame (US Statistical Policy, 2001). Three main types of error are well distinguished with regard to this type of nonsampling error. *Undercoverage* - there are target population units that are not accessible via the frame. *Overcoverage* **-** there are units accessible via the frame, which do not belong to the target population. And *multiple listings* (*duplications*) are called the target population units presented more than once in the frame (EHQR, 2009).

Overcoverage can be detected during the measurement process and is straight forward to handle in the estimation procedure. It results in increasing sampling error and survey costs. Multiple listings can also be handled by researchers and also result in an increase of sampling error and cost, but no significant bias. *Comparison of the sampling frame with the frame*

[1] TSE is the cumulative effect that all sources of error in a survey (sampling and nonsampling) have on the distribution of the estimates (US Statistical Policy, 2001). Sampling error applies only to sample surveys and is due to the fact that only a sample is surveyed rather than all the units in the population. The nonsampling error does not involve the type of sample that we draw, but rather everything else in the survey production process, being the interviewers, the respondents, the questionnaire, nonresponse, the general population list that we use to draw the sample (sampling frame).

population and *matching of the interviewed units with the units in the population frame* are among the methods used to control the overcoverage and multiple listings (Atanasov, 1990).

Quantitative information on overcoverage and multiple listings is normally easy to obtain in sample surveys. Thus the attention of survey methodologists is more frequently focused on undercoverage. Unlike the other types of coverage error, undercoverage appears to face serious problem in terms of measurement. Undercoverage cannot be detected in the measurement process and its rate can only be evaluated through control checks that are external to data collection. That is why it is perceived as the most serious type of coverage error, which results in difficult to detect and evaluate biases. The resulting bias depends on the units outside of the frame population, but in the target population and the differences between the characteristics of these units and those in the frame population. Thus a first step in the assessment of the undercoverage bias is the qualitative description of these units.

Two methods are mainly used to detect undercoverage and to assess its magnitude (EHQR, 2009). One of them is the *analysis of lag structure*, applied when there is a time lag in registering frame units and the later version of the frame is perceived as the one that can provide additional updated information. Every frame is updated with a certain lag (e.g. birth, death, or change of a unit is registered with a delay). Due to this the frame will always, to a smaller or larger extent, have less than perfect coverage at the time of use. The lag effect can be studied for example by matching two consecutive register versions and establishing which of the units in the latter version should, by definition, have been included in the former. Other approaches are also possible. Register errors can be studied in several consecutive versions. It may be possible to observe certain stability in error levels that can be assumed to continue into the future. The degree of undercoverage (or overcoverage) can thereby be estimated.

Comparison with another frame or other external information (e.g. matching with a different register), is the second method to detect and try to assess undercoverage. When applying this method the sampling frame is matched with a control register that entirely or partly covers the same population as the frame. Matching can be also done on a sample basis. If the control register is of superior quality, then errors in the frame can be directly assessed. Otherwise a reconciliation process, involving checking (a sample of) the non-matches is needed to determine the extent of errors in the survey frame (EHQR, ESQR, 2009; US Statistical Policy, 2001; Atanasov, 1990).

However, the methodological literature dealing with the problems of coverage error more frequently focuses on the assessment on the number of missing or erroneously included units and their characteristics than their effect on the survey results. The assessment of the bias as a result of coverage imperfections is a not always feasible and easy to be done survey task. The availability of relevant external sources of information to make possible the assessment of the magnitude of the coverage bias is infrequent.

2.2 Measuring nonresponse

Nonresponse is the failure of a sample survey to collect data for all data items in the survey questionnaire from all the units designated for data collection. In other words - the difference between the statistics computed from the collected data and those that would be computed if there were no missing values is the nonresponse error (EHQR, 2009). Like the coverage error, the nonnresponse error is an error of nonobservation. However, nonresponse error differs from coverage error in that nonresponse reflects an unsuccessful attempt to obtain the needed information from an eligible unit, whereas coverage error reflects the failure to have the sample unit uniquely included in the frame (US Statistical policy, 2001).

There are two types of nonresponse error – *unit nonresponse* and *item nonresponse*. *Unit nonresponse* occurs when no data are collected about a unit designated for data collection, and *item nonresponse* occurs when data only on some but not all the survey data items are collected about a designated unit. The extent of response and accordingly nonresponse is measured in terms of response of two kinds – unit and item response rate. Unit response rate is the ratio of the number of units for which data for at least some variables have been collected to the total number of units designated for data collection[2]. Item response rate is the ratio of the number of units which have provided data for a given data item to the number of units that have provided data at least for some data items. However, no rules aiming to provide a clear distinction between item and unit nonresponse (or in other words – when partly completed interview is treated as unit nonresponse) have been introduced yet, therefore researchers follow different practices. There are also other ratios used to adjust the survey output in terms of survey procedures imperfections, where the purpose is mainly to increase the weights of the respondent cases to represent the nonrespondents. For example the design-weighted response rate, that sum the weights of the responding cases according to the sample design, and the size-weighted response rate, that sum the values of auxiliary variables multiplied with the design weights, instead of the design weights alone (EHQR, 2009)[3].

It is often assumed (correctly or not) that the lower the response rate, the more question there is about the validity of the sample (AAPOR, 2011). However, the response rates provide only indirect indications of the bias risks and the actual bias depends mainly on the relative differences between the respondents and nonrespondents with respect to a survey variable. But although response rate information is not sufficient for determining how much nonresponse error exists in a survey, or even whether it exists, calculating the response rates is a critical first step in understanding the presence of this component of potential survey error[4].

As the actual degree of bias due to nonresponse is a function not only of nonresponse rate, but also how the characteristics of respondents and nonrespondents differ, the effects of nonresponse error cannot be directly observed and evaluated due to the difficulty to collect data from and for nonrespondents[5]. Although the difficulty in measuring nonresponse error provokes researchers mainly to search ways to limit and minimize sources of nonresponse, there are a few specific methods aiming to evaluate the magnitude of this type of error.

[2]Calculations could be done on both unweighted and weighted rates of response. Each calculation could be useful for particular aim of the survey. Unweighted response rates could be useful quality indicator for the process of data collection and could be calculated on national, regional or also at interviewer level, or at the level of definite administrative areas in order to assess the performance of interviewers and supervisors. Unweighted nonresponse rates could be used as an indicator of interviewers efforts in the data collection process. Weighted response rates could be useful for the overall assessment of nonresponse effect on survey data (Kasprzyk & Kalton, 1997; Madow, Nisselson, & Olkin, 1983). Unweighted response rate is frequently used for monitoring of the fieldwork progress (US Statistical policy, 2001).

[3]Although response rate seems relatively easy to be calculated, it could be considered as a problem that the different research organizations apply different methods in its calculation. Thus the efforts of some international research organizations are focused in the standardization of methods to calculate response rates (CASRO, 1982; AAPOR, 2000) which however have no total success yet. Nevertheless, AAPOR is continuously trying to establish main working frames in calculating response rate that could be applied to different survey modes (AAPOR, 2011).

[4]It may be also useful to calculate other response rates for specific purposes, for example rates of completion, rates of refusals or noncontacts, contact or cooperation rates, response rates at various levels to monitor data collection and interviewers performance.

[5]There are attempts to gather data on nonrespondents and noncontacted units through paradata, however this issue also provokes ethical related and organizational discussions.

The basic approach is to compare the response and nonresponse strata with regard to any variables that are available for both these strata. According to *Lessler* and *Kalsbeek* (1992) the potential of the systematic error due to nonresponse could be evaluated through *identification studies* (US Statistical policy, 2001). These types of studies compare the characteristics of the responded units with those of nonrespondents on variety socio-demographic characteristics available through the sampling frame or external sources. If the distributions of respondents are similar to those of nonrespondents on the available variables, the bias concern should decrease. On the contrary, if the distributions differ, the observed variation could provide some insights for the existing differences. Variation of the identification study is the studying of respondents` characteristics during the different editions of the survey in longitudinal surveys. Another variation is the comparison of respondents and norespondents in sample survey with sources of exhaustive information (e.g. census). The identification studies could be linked with the *Completing with register data* method frequently applied by the European Statistics (EHQR, 2009). *Completing with register data* is a method assuming that there is a strong enough correlation between a survey variable for which there is nonresponse and another variable in the sampling frame or another register. This information can be utilized in different ways. For evaluation, one way is to compare the estimate of this other variable from the whole sample with that derived from the sample excluding nonrespondents. A small difference provides some indication of a small nonresponse bias for the survey variable as well. The better the correlation is between the two variables, the better is the judgement that can be made in this way.

There are also *special data collection* methods aiming to show how the nonresponse error would change if higher response rate were achieved. These studies are applied so that a higher response level is reached than the one achieved with a normal effort. For example, more effort can be done for tracing, persuading (soft) refusals, increasing time for field work, allowing other data collection modes, reducing response burden by concentration on fewer variables or by offering incentives to respondents. The differences in estimates thus obtained will reflect not only nonresponse error, but also measurement and random sampling errors.

Another technique (applicable in longitudinal surveys) is to study the *variations over response waves*. The purpose of this approach is to show how estimates were changed as a larger share of data collection is accomplished. Results are of interest when intending to publish flash estimates based on data obtained before a certain data. Another use arises in the context of budgetary or timeliness purposes, to reduce the target response rates and to be able to judge in advance the consequences of such a reduction. A more controversial use of such studies is to draw conclusions about the remaining nonrespondents based on those that responded in the last wave (EHQR, 2009).

Taking into account that the problems related to the complete (unit nonresponse) or partial (item nonresponse) loss of data could not be entirely avoided in the research practice, survey methodologists have to be prepared how to compensate nonresponse (to optimize the data loss). There are a few techniques, including the choice to take no measures or make as much effort to increase the number of respondents (Lessler and Kalsbeek, 1992; Stoop, 2005; Markova, 2004). Two general post-fieldwork approaches in compensating nonresponse (missing data) are well known and relatively frequently applied in the practice – *adjustment* as part of the estimation process (e.g. *weighting adjustments*) or direct estimation of the values each nonrespondent might have reported if he/she has been a respondent (*imputation*). Kalton and Kasprzyk (1986) indicate that corrections through weighting are more proper in compensating missing data at unit level, while the procedures of imputation are more appropriate in terms of item nonresponse (US Statistical Policy, 2001).

2.3 Evaluating Measurement Error

Measurement error occurs as part of the data collection, and as it is related to the observation of the variables being measured in the survey, it is also referred to as observation error. It may arise from four main sources: the questionnaire (e.g. wording, design, indicators operationalisation, etc.), the method of data collection or the effect of the mode of administration (e.g. in person, telephone, mail, etc.) on respondent answers, the interviewer effect on the respondent, and the respondent himself on the base of different experience, knowledge, attitudes (Biemer et al., 1991). The setting/place, conditions of interviewing is also classified by some authors as a source of measurement error (Atanasov, 1990), but also the third party presence, the ability of the respondent to completely understand the questions due to language barrier, mental problems, etc.

Measurement error can be characterized as the difference between the value of a variable provided by the respondent and the true (but unknown) value of that variable. The total survey error of a statistic with measurement error has both fixed errors (bias) and variable errors (variance) over repeated trial of the survey. Measurement bias or response bias reflects a systematic pattern or direction in the difference between respondents` answer to a question and the correct answer. Simple response variance reflects the random variation in the respondents` answer to a survey question over repeated questioning (US Statistical Policy, 2001).

When the risk of substantial bias is considered high, evaluation studies are needed. Respondent error can be assessed by a *re-interview study* in which the respondent is asked to provide the same data on a second occasion. If there is no memory effect, the two interviews may be considered independent and the difference between the responses should be regarded as an indication of the size of the measurement error.

In order to assess instrument or interviewer effects, repeated measurements can be made with different instruments, e.g. *alternative phrasing of questions or different interviewers*. Alternatively, an *experiment* can be carried out *with subsamples being randomly allocated to different instruments and/or interviewers*. This approach is mostly appropriate for surveys on attitudes/opinions or where memory effects are involved.

Four groups of methods are applicable for evaluating errors at unit level (EHQR, 2009). Such errors could have been generated in the measurement phase, the processing phase or they could have existed already in the sampling frame.

The *comparison with other information at the unit level* is considered the best way to obtain a quality check provided there is a common unit identification scheme for both sources. Matching of registers, as mentioned under the coverage errors above, can be used also for this purpose, provided the control register can be assumed to have good information about the units for certain variables. Care must be taken to distinguish actual errors from differences in definition or measurement points in time.

Another useful approach is the *re-interview with superior method (*or *control at source)*. Control at source means that the evaluator gets access to source data – company accounts or records kept at an agency, etc. A re-interview with a superior method may use an expert interview or face-to-face instead of mail interview. Another approach is to use the same interview method once again but with a different interviewer and use a reconciliation procedure (for example an expert panel) where different responses are obtained. Such methods capture all types of error that have occurred during measurement and processing, whether due to respondent, questionnaire, interviewer or data entry. They are best done by means of a random sample of units resulting in unbiased estimates of error.

Replication as a method measuring the error of observation means that there are two or more observed values for a sampled survey unit. Such values can be obtained by different interviewers, from different respondents (answering for the same sampled unit) or simply by repeating the measurements after sufficient time so that the respondents will not remember their initial responses. The differences between the measurement values can be used for learning how stable the measurement process is. Formal analyses of replication often assume that errors are independent between replications. This assumption is rarely fully met in practice. The method is used for estimating the random variation due to measurement. Under some circumstances (for example if an expert interviewer or respondent is used) it can also provide some information on the systematic error (bias).

By comparing results from original and edited data the extent of initial measurement error can be deducted. Studying the *effects of data editing* gives a minimum estimate of the error levels, since not all errors will be detected in the editing process. Such analyses provide ideas for improving the measurement methods, but no information on the undetected measurement errors, nor how they affect the statistical outputs.

2.4 Measuring Processing Error

Data processing comprises multiple steps each generating errors due to different sources – from the more simple errors of data entry (including errors due to transcription or data transfer) to the more complex errors arising from the erroneous specification of the coding, editing (checks and corrections) or imputation models. Processing errors affecting individual observation cause bias and variation in the survey output, just as measurement errors do.

The evaluation of data processing is especially important when the coding procedure of response data is done manually. The quality of coding operations depends in a complex way on the coding rules, how they are interpreted in practice and on downright mistakes committed by the coders (ESQR, 2009).

Two procedures are frequently used in studying the effects of editing and coding (EHQR, 2009). The *effects of editing* are obtained by comparing edited and unedited data. By calculating the final estimates based on both data sets, the total net effect of editing can be measured. These effects can be broken down by unit in a so called top-down list, where the effects by unit are sorted in descending sequence and the most influential units can be seen. Such a list can serve several purposes. One is to check once more that the influential units have their correct values another is to generate ideas for optimizing the editing procedures. The *study on coding variation* could be done in two variations. In the *independent coding control study* the coding is done twice without the coders being allowed to see each other`s results. In dependent coding the second coder has access to the first coder`s proposals. *Dependents coding* gives, as expected, smaller variation between the coders. High coding variation is an indicator of a large potential processing error.

Discussion

Nonsampling error and its specific components remain difficult to be measured in practice. As demonstrated in the article, approaches towards the evaluation of each specific type of nonsampling error exist. And even if they are applied in practice, the outcome of the assessment rarely shows quantitatively the magnitude of the error effect. The capabilities of most of the above enumerated methods to produce quantitative estimates are limited. Response rate is among the most frequently calculated quantitative indicators of data quality, however the different practices of its calculation (and also the missing methodological rules providing clear distinction between item and unit nonresponse, and the treatment of partly completed interviews) raise the awareness of response rates comparability within surveys. Moreover, this indicator itself does not provide direct information on the error of nonresponse. Studying the characteristics of nonrespondents, but also studying the characteristics of omitted in the sample target units, always needs further research efforts.

Although still a challenge to survey researchers, the attempts towards nonsampling error decomposition and efforts to study and evaluate the effect of specific types of error place researchers aspiration to survey quality improvement a step forward.

References:

American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th edition. AAPOR.

Atanasov, A. (1990). *Accuracy and Control in the Sociological Survey*. Marin Drinov Edition. Sofia.

Biemer, P. (2010). *Overview of Design Issues: Total Survey Error*, Handbook of Survey Research.

Biemer, P. (2010), *Total Survey Error:* Design, *Implementation, and Evaluation,* Public Opinion Quarterly, Vol 74(5).

Beimer, P., Lyberg, L. (2008), *Quality Assurance and Quality Control is Surveys*, International Handbook of Survey Methodology.

Biemer, P., Lyberg, L. (2003), *Introduction to Survey Quality*, Wiley Series in Survey Methodology. Hoboken, NJ: John Wiley & Sons, Inc.

Czaja, R.F., Blair, J. (2005), *Designing Surveys*: *a guide to decisions and procedures*, 2nd ed. London: Sage.

EUROSTAT. (2009). *ESS Handbook for Quality Reports*. Methodological and Working Papers.

EUROSTAT. (2009). *ESS Standard for Quality Reports*. Methodological and Working Papers.

Executive Office of the President of the United States. (2001). *Measuring and Reporting Sources of Error in Surveys*. Statistical Policy: Working paper 31.

Groves, R., Lyberg, L. (2010), *Total Survey Error: Past, Present and Future,* Public Opinion Quarterly, Vol 74(5).

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

Loosveldt, G., Catrton, A., Billiet, J. (2004), *Assessment of Survey Data Quality: A Pragmatic Approach Focused on Interviewer Tasks*, International Journal of Market Research, Vol. 46(1).

Markova, E. (2004). *Problems of Sample Optimization in Case of Missing Data in Surveys*. Ph.D. Thesys, Institute of Sociology, BAS.

Stoop, I. (2005). *The Hunt for the Last Respondent*. The Hague: Social and Cultural Planning Office of the Netherlands.