Approaching Quality in Survey Research: Towards a Comprehensive Perspective*

JINDŘICH KREJČÍ**
Institute of Sociology AS CR

Abstract: The article has two goals: (1) bring attention to the problem of inappropriate treatment of survey data quality issues in the social sciences, and (2) introduce the basic principles of contemporary approaches to survey quality. If quality evaluation focuses solely on sampling error, most aspects of data quality are ignored and surveys are assumed to have 'ideal' statistical characteristics that are rarely attainable in the pragmatic world of survey fieldwork. A complex overview of the entire process of data collection provides a more solid foundation for evaluating data quality. Under this approach, quality is ensured by controlling the whole survey process. Accuracy, which is commonly elaborated using the concept of survey error, ceases to be the only dimension of quality. Nevertheless, this data quality component is crucial for data analysis and statistical testing. A comprehensive approach to survey data quality requires us to take account of complex sample designs when evaluating sampling error and to identify and distinguish between different dimensions of nonsampling error. Analysts who are not directly involved in data collection have limited ability to obtain information necessary for data quality evaluation. There are two types of quality standards: administratively imposed standards (ISO20252:2006) and the technical and ethical criteria of professional associations (e.g. ICC/ESOMAR, AAPOR/WAPOR, SIMAR). These help break this information barrier between data producers and data users.

Keywords: survey research methods, survey accuracy, total survey error, quality standards
Sociologický časopis/Czech Sociological Review, 2010, Vol. 46, No. 5: 1011–1033

Undoubtedly, the quality of data determines the ability of empirical social research to provide a valid and reliable explanation or description of social reality. Notwithstanding this fact, quantitative social science studies are often based

* Work on this article was supported by the grant 'COMPDAT – Data Sources, Research on Data Quality, Standards, and Methods of Data Harmonisation for the Purposes of International Comparative Social Research and Integration into the CESSDA Network' (Reg. No. LA09010). This project is financed by the Ministry of Education, Youth and Sports of the Czech Republic. The author would like to thank Pat Lyons for very useful recommendations and his help with the English version of the paper and two anonymous reviewers for their valuable comments.

** Direct all correspondence to: Jindřich Krejčí, Institute of Sociology, Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Prague 1, e-mail: jindrich.krejci@soc.cas.cz.

on incorrect assumptions concerning the quality of the data analysed. A great portion of such data currently originates in sample surveys. Methods of statistical analysis are frequently used without due attention to the limitations of this type of data and assumptions concerning the processing of survey data. Statistical inference on societies is routinely based on estimates of sampling error that are valid for simple random samples, while general population samples of such quality are extremely rare.

These and many other inappropriate assumptions about survey data often pass through the peer-review processes of leading social science periodicals without any comments. While social science researchers silently consider their methods as 'probably' very robust and do not undertake detailed debates on the quality of their data, experts from other research areas and sometimes also the non-professional public look at their results with growing scepticism.

The reason for many problems are the widespread selective approaches to data quality, which respect only the basic roots of the methods in mathematical statistics and psychometrics, but do not consider the specific nature of how surveys are organised in reality. Many researchers do not have sufficient knowledge of complex surveying methods and simply do not understand survey data quality issues. Consequently, they do not pay attention to a number of important indicators reflecting the character and possible limitations of the data processed in their analysis, produce erroneous interpretations, and are not able to justify their procedures if the quality of the survey data is criticised.

This article has two goals.[1] First, it aims to draw attention to the problem of the inappropriate treatment of survey data quality issues in social sciences and to the origins of this problem and its impact on social research. Second, the article will introduce the basic principles of contemporary approaches to survey quality. The final section will outline differences in social survey quality standards.

Social survey methods and traditional approaches to survey quality

This article concentrates on quality issues and is specifically concerned with social survey research. This kind of research is based primarily on the analysis of data from sample questionnaire surveys, which is generally understood to be a method that consists of conducting standard structured interviews on a representative sample of units (individuals, households, organizations, etc.) selected from a target population for the purpose of obtaining aggregate information and inferring the characteristics of the target population. In this respect, the term 'representativeness' means the quality of survey sampling and other methodological

[1] This article is based on an extension of the introductory sections of some of the author's presentations of survey quality issues published in Czech in a monograph entitled 'The Quality of Social Science Sample Surveys in the Czech Republic' (*Kvalita sociálněvědních výběrových šetření v České republice,* Prague: SLON 2008).

procedures which ensure that the sample of units interviewed has a distribution of characteristics that may be deemed equal to that of the population sampled [Babbie 2004; Kish 1965; Groves et al. 2004].

The basics of the methods and principles of sampling for the purposes of reasoning about the whole population come from the study of probability theory within mathematics. In line with these principles, the quality of sample surveys has been traditionally viewed as a function of measurement accuracy or its complement survey error. A typical understanding of the survey sampling process is based on a distinction between sampling error (sometimes referred to as 'statistical error') and nonsampling error. Such a distinction reflects the statistical method's theoretical assumptions and character. Moreover, use of probability theory makes it possible to calculate estimates of measurement uncertainty such as sampling error.

Survey sampling has its roots in probability theory, but it also draws inspiration from psychometrics and its use of interview techniques to gather data.[2] In this respect, the concepts of validity and reliability and their measurement have been developed with the goal of constructing instruments such as questionnaire items and scales. In other words, measurement concepts are evaluated on the basis of how well they correspond to the real world and the extent to which alternative and repeated survey measures yield consistent results.

Even if the above-described approaches reflect the basic principles of the survey method, they do not fully cover the issue of survey research quality. The implementation of a representative survey does not include only the construction of a sample from a target population and creating a questionnaire. In the real world it is a complex process that consists of a series of interrelated tasks of a different character. Traditional approaches to survey quality that only consider key principles stemming from mathematical statistics and psychometrics embrace a small portion of the total set of survey quality issues, and are thus insufficient foundations for survey research.

The different assumptions of data analysts and data collectors

Hubert M. Blalock Jr. [1989], in an essay on the contributions of quantitative sociology, noted two contrasting trends. On the one hand, there are improvements in the quality of data analysis stemming from the computer revolution and the increasing availability of large datasets. On the other hand, he noted persistent problems in the quality of data collection, conceptualisation, and measurement evident in many research areas. Some of these problems arose as a by-product of the necessary process of specialisation in the social sciences, where gaps in communication resulted in some constraints on the contribution that quantitative sociology was

[2] Psychometrics is primarily associated with the theory and measurement of attitudes, beliefs, personality traits and educational achievement.

making to the study of society. In this respect, Blalock [1989: 450] also noted the widespread habit of applying sophisticated analyses to poor quality data.

This problem persists to this day, some twenty years after Blalock's [1989] seminal article. The situation regarding data quality in sample surveys can be taken as a typical example of the communication gap. The time is long gone when a single team routinely implemented the research process from start to finish. Researchers are specialised and only have control over certain stages in the research process. These days, fieldwork on data collection is often subcontracted to specialised agencies and researchers are now able to download an increasing number of datasets from web-based data archives. Increasingly, the practices of data collection and data analysis are becoming separated.

In a slightly different context, Robert Groves [2004: 1–37] has demonstrated the significant differences between the ways survey errors are treated by those who collect data and those who analyse it. Data collectors need to prevent problems, work systematically on removing them, and aim at continuously producing reliable results, while analysts need to assess the accuracy of the specific data they work with and estimate the error of specific statistics for the purposes of statistical testing.

Thus, many analysts do not come into direct contact with the survey organisation and the fieldwork, are little interested in these stages of the research process, and often have little understanding of the data generation process. Often they are satisfied with an assurance that the survey data are representative, or with a short outline of the sampling method used to create the dataset. Even if data quality is evaluated, analysts set their priorities primarily in terms of the formal requirements of statistical testing.

However, in the practice of social research, the phenomenon of data quality does not exist separately from data collection. If quality evaluation only focuses on sampling parameters, this leads to two important consequences: (1) most aspects of data quality are ignored, and (2) the survey data are assumed to have 'ideal' statistical characteristics that are rarely attainable in the pragmatic world of survey fieldwork. In such situations, sample surveys often turn out to be a kind of black box whose details are largely unknown to the data analysts. Thus, data analysts lose control over their research and can only hope for favourable statistical test results.

This 'division of labour' also results in an unfortunate process of negative feedback. Since analysts use the products of data collection agencies and sometimes are their direct clients, the unrealistic expectations of analysts have a counter-productive effect on the performance of survey fieldwork agencies and hence the general level of data quality within the survey data market. A survey agency's efforts are necessarily focused on the priorities of clients, and these efforts may be made at the expense of investments into the less demanded but more important parameters of quality. Thus, in a competitive market, the client's ability to recognise quality is always an important precondition for investment into more valuable production.

Overemphasising statistical significance

In a recent article Petr Soukup and Ladislav Rabušic [2007] criticised social researchers' 'obsession with statistical significance' and the often inappropriate use of statistical techniques for making causal inferences [see also Gill 1999; Denis 2003; Gigerenzer, Krauss and Vitouch 2004]. These authors demonstrate that statistical significance has become the 'gold standard', justifying all conclusions, and is frequently applied in contradiction of key statistical principles. Soukup and Rabušic [2007] have noted that some of the fundamental theorems within mathematical statistics place important limitations on the application of statistical techniques within social research.[3]

In this article, attention is focused on the other side of the coin. The use of survey data is not only constrained by statistical considerations, it is also fundamentally determined by how surveys are actually created during fieldwork. The study of issues of measurement accuracy, total error, estimation bias, and measurement variance have long rested in the hands of mathematical statistics, and that is why the most work related to surveying has focused on sampling error. Nonsampling errors, which are measured with more difficulty and less reliability, have often been treated as unobserved variables of marginal importance or conceptualised in vague and selective ways.

However, mathematical probability theory primarily works with an entirely different kind of data than, for instance, sociology. In social research, we do not draw assorted coloured balls from a hat, which is a relatively easy thing to do. Instead, highly complex sample surveys must be organised in order to collect data. Each respondent must not only be selected but also found, contacted, and persuaded to participate in the survey. Finally, the required data must be collected appropriately from the respondent. The ways we manage such a complex process have important consequences for several types of errors in the resulting dataset.

Evaluations of data quality that are based solely on the principles of mathematical statistics build on a number of hypothetical assumptions that are usually not fulfilled in practice. This is because they cannot be fulfilled in real-life social research. As a result, there are analyses that completely ignore the real character of the survey data under examination. Moreover, unrealistic assumptions about survey data are often a justification for detailed and expensive mathematical examinations of partial errors which play only a minor role in the total error.

[3] It should be noted that much of this criticism refers to conventional social and behavioural science practices using classical statistical techniques. Within the past decade greater interest in reporting effect size, confidence intervals, Leamer bounds, and the use of power analysis, meta-analysis, predicted values, simulations, or employing a Bayesian modelling framework have resulted in increased understanding of the limits of null hypothesis significance testing [Gill 1999; Denis 2003].

Turning attention to nonsampling errors

The dangers of overemphasising sampling error have been known for a long time. Experienced researchers have always strived for a balanced allocation of survey resources in order to prevent nonsampling errors as well [see Bell 1991; Noelle 1963]. A part of their effort has also been analysed and documented in methodological literature [e.g. Merton and Lazarsfeld 1950]. According to Biemer and Lyberg [2003: 8–11, 305–350], the issues of sampling error began to be implemented in the methodologies of social surveys in the 1930s, along with the introduction of contemporary sampling methods. However, as early as the 1940s it became generally known that sampling error does not fully account for total survey error. A systematic scholarly debate was stimulated by the so-called 'US Census Bureau Survey Model' formulated by Hansen and his colleagues [see, e.g., Hansen, Hurwitz and Pritzker 1967]. Using Census Bureau surveys this model demonstrated that sampling variance only captures one type of error, and crucially total error will be underestimated if it is attributed solely to sampling error.

A gradual process of introducing formalised rules and comprehensive approaches to nonsampling errors began in the 1970s. Based on the above-mentioned US Census Bureau Survey Model, procedures for estimating nonsampling errors were developed by Hartley and Rao, Bailar and Dalenius, and other scholars [Biemer and Lyberg 2003: 8–11; Groves 2004: 15; Bell 1991; Dalenius 1977]. In the early 1980s, Dalenius [1981] proposed an integrated approach to survey organisation, resulting in a concept entitled 'Total Survey Design'. Over the past twenty years, we have seen a major reconsideration of the rules for evaluating and managing quality in research organisations. National statistical bureaus have been the pioneers of this change [see Morganstein and Marker 1997].

Among the most important and comprehensive works that have formed the basis for the contemporary theoretical and analytical literature on the topic are Andersen, Kasper and Frankel's [1979] *Total Survey Error*; Lessler and Kalsbeek's [1992] *Nonsampling Errors in Surveys*, and Groves' [2004] *Survey Errors and Survey Costs*. Systematic textbook-like introductions to the subject of survey data quality have been written and published by Biemer and Lyberg [2003] and Groves et al. [2004]. The explicit specification of quality standards in survey research for one of Germany's main research funding organisations, Deutsche Forschungsgemeinschaft (DFG), is seen as an important step in the European context [Kaase 1999].

A comprehensive approach to survey quality

Sample size and sampling method can be viewed as the basic parameters of survey quality, since they determine if the method of representative survey was used or not. However, they do not fully cover either the issue of data quality or accuracy. Concepts of quality focusing predominantly on sampling error do not allow researchers to control the total error in designing and realising surveys and fail to provide a sufficient basis for making evaluations of survey quality. Therefore, they are also untenable in light of the principles of the contemporary 'quality revolution', which is based on the ideas that (1) problems must be prevented, rather than solved once they arise, and (2) quality must be monitored and improved continuously. Moreover, from this perspective, the quality of the product is ensured by controlling the quality of the production process.

Juran and Gryna [1993] define quality as 'fitness to use'. However simple this sounds, it provides a point of departure for a comprehensive approach to quality. Sample surveys are no exception in this respect. Statistical accuracy ceases to be the only or central dimension of data quality, as other equally important dimensions emerge. Biemer and Lyberg [2003: 14] have summarised the reasons for such a shift as follows:

1. Accuracy is difficult and expensive to measure, so much so that it is rarely done in most surveys, at least not on a regular basis. Accuracy is usually defined in terms of total survey error; however, some error sources are impossible to measure. Instead, one has to assure quality by using dependable processes, processes that lead to good product characteristics. The basic thought is that product quality is achieved through process quality.
2. The value of post-survey measures of total survey error is relatively limited. Except for repeated surveys, accuracy estimates have relatively small effects on quality improvement.
3. The mechanical quality control of survey operations such as coding and keying does not easily lend itself to continuous improvement. Rather, it must be complemented with feedback and learning where the survey workers themselves are part of an improvement process.
4. A concentration on estimating accuracy usually leaves little room for developing design quality components.

On this basis, numerous research organisations have gradually reconsidered their definitions of quality. This began with the criteria used by statistical bureaus, and new approaches have gradually prevailed in the practices adopted by survey research organisations in both the public and private sectors alike.

For example, different statistical offices currently apply sample survey standards that are very similar in their contents [see Biemer and Lyberg 2003: 13–25; Lynn 2001]. As a rule, between six and eight dimensions of quality assurance and control are prescribed. On this basis, models of Total Quality Management (TQM)[4] have been defined which set the required characteristics of a final product, de-

[4] TQM – Total Quality Management represents a business management philosophy which concentrates on the entirety of product or service quality and a continuous improvement thereof. It has served as a basis for defining numerous general TQM models, which are further specified and adopted for the circumstances of individual organisations applying them.

fine partial goals, elaborate the individual dependable processes to achieve them, identify and treat problematic points, and specify control points, procedures of quality monitoring, learning processes, and feedbacks for ensuring continuous improvement [see Morganstein and Marker 1997; Colledge and March 1997].

As an example, we will describe the TQM model of the European Statistical System (ESS), which was formulated by Eurostat in order to manage numerous activities of European national statistical bureaus and other institutions supplying data to the System [Eurostat 2002, 2009]. The underlying concept of survey quality is just a little different from the general ISO definition of quality as applied in business and production: quality is defined as 'the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs' [ISO 1994]. More specifically, Eurostat determines the quality of ESS statistics on six dimensions [Eurostat 2003, 2009]:

- Relevance: how statistics meet current and potential users' needs;
- Accuracy: the closeness of computations or estimates to the true values;
- Timeliness and punctuality: the time lag between the event and the availability of information about it; and the time lag between the release of data and the target date when it is needed;
- Accessibility and clarity: the conditions in which users can obtain data; and the level of documentation and accompanying information, including information on data quality and availability of professional assistance;
- Comparability: the impact of applied concepts and measurement tools on comparability over time, between geographical areas or between subject domains, i.e. the extent to which differences between statistics reflect real differences between areas on the three dimensions;
- Coherence: the adequacy and trustworthiness of different combinations of statistics and the various uses of thereof.

The aim of this TQM model is to provide a framework for achieving quality in the production of statistics and promote the harmonisation of the quality of statistical processes and outputs within the ESS. Like other models, TQM does not guarantee a high level of quality of each final product, but it affirms that there is a systematic effort in achieving quality, including controls and reporting and structures able to follow the rules. The availability of quality reports and quality and performance indicators can also help researchers to properly use obtained statistics in their analysis.

Similar dimensions are defined by other national statistical offices and are used for international statistical services and many other research institutions.[5] It should be added that these principles may be applied quite differently in research practice. For instance, while the International Monetary Fund (IMF) uses similar

[5] For instance, Statistics Canada uses relevance, accuracy, timeliness, accessibility, interpretability, and coherence [Statistics Canada 2002]; Statistics Sweden uses the following

dimensions as Eurostat, it places much more emphasis on qualitative indicators in the evaluation process [Laliberté, Grünewald and Probst 2004].

Public opinion polling and market research agencies as well as smaller academic research institutions conducting sample surveys often do not have such highly elaborated rules. However, they usually rely on similar principles. The influence of contemporary approaches to sample survey quality is apparent in the standards and guidelines of international research associations such as the ESOMAR (World Association for Opinion and Marketing Research Professionals) or the WAPOR (World Association for Public Opinion Research), as well as national associations, including the Czech SIMAR (Association of Market Research Agencies). The above-described principles also provide a basis for a specialised standard, the ISO 20252:2006, for market, opinion, and social research.

There is one basic quality criterion that is not mentioned by Eurostat as an independent dimension, yet it is taken into account by all the other dimensions of its system of classifying sources of statistical errors: survey efficiency. The importance of the relationship between costs and accuracy in sample surveys was demonstrated persuasively by Groves [2004] in his book, *Survey Errors and Survey Costs*. First published in 1989, this book serves as one of the fundamental textbooks on survey errors. Indeed, saving money is one of the important motivations for conducting sample surveys. Taken to the extreme, we would only conduct census surveys, rather than sample surveys, if costs (financial and other) did not matter and accuracy was the only criterion.

In the real world of limited budgets survey costs are a very important consideration in all research work. However, once we put straight the fact that 'price' matters, we are obliged to assess the cost of applying each method in advance. When choosing between different methods, when adapting our methodology in some way, we must always ask the following two questions. First, what level of accuracy is necessary for the given purpose, and what resources are available? Second, how exactly will accuracy improve if the particular adaptation of methodology is implemented, and at what price will the improvement occur? In other words, one must determine if the accuracy improvement is worth the money, effort, and time involved. One must determine how well a specific strategy performs when contrasted with alternative procedures. Such questions should be asked for each method applied and for the survey methodology as a whole. Therefore, in sample surveys, maximising on survey accuracy does not mean striving for the best quality possible. Instead, the consumers of survey fieldwork aspire for an acceptable level of accuracy that can be supplied at an acceptable price.

The relationship between cost and accuracy is not the only dilemma affecting quality criteria. Other typical dilemmas include those of accuracy versus the timeliness of delivering results, relevance versus comparability over time, rel-

quality dimensions: content, accuracy, timeliness, comparability/coherence, and availability/clarity [Biemer and Lyberg 2003]; OECD uses relevance, credibility, accuracy, timeliness, accessibility, interpretability, and coherence [OECD 2002].

evance versus accuracy, etc. The quality of a research survey is therefore not limited to achieving favourable indicators for the different criteria. Instead, one must carefully consider multiple criteria and partial research goals in order to achieve an ideal compromise between many variables.

Survey error

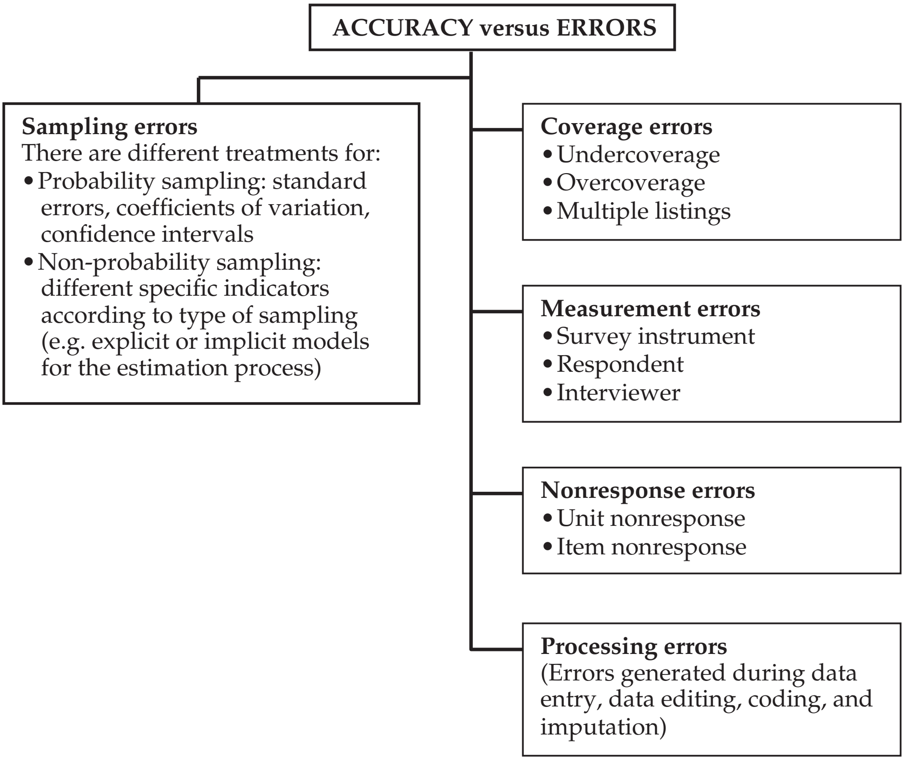
As described above, accuracy is not the single dimension of data quality. However, this quality component is crucial for the verification of results of statistical analysis and especially for purposes of statistical testing. That is one reason why accuracy is the subject of strong interest from data analysts, and it will be discussed in more detail below.

In sample surveys, accuracy represents an estimated level of difference between the survey estimate of an attribute and its true value. It is the opposite of error and often can be better understood by using the concept of survey error. The foregoing discussion has made it clear that reliable information on survey data quality cannot be obtained if survey error is conceptualised as sampling error plus an unobserved, undifferentiated, residual nonsampling error. The unobserved residue is too high to ignore. A comprehensive approach to nonsampling error requires us to identify and distinguish between different dimensions of nonsampling error.

The work of Robert Groves [2004] provides a foundation for much methodological and analytical literature on survey data quality in contemporary social research. Therefore, we will use his classification of total error and its components as a point of departure for our introduction into the complexity of the problem of survey accuracy. Groves [ibid.] builds on the works of Kish [1965] and Andersen, Kasper and Frankel [1979]. In his view, total survey error is just a theoretical construct. Errors are specific for individual survey statistics, rather than the survey as a whole. Total survey error is thus understood as the mean square error. At the same time, it is the sum of all variable errors (arising out of measurement variance) and all biases (differences between measured values and true values). Bias is a constant error affecting a statistical estimator in all applications of a given survey design, i.e. the same error always arises when the same design is used in surveys. Variable error arises from the fact that measured values differ across units that are the sources of error. Both types of errors are interconnected and can only be distinguished if the survey is repeated.

Furthermore, Groves [2004] provides a classification of errors according to their sources. There are two types of errors. Errors of non-observation arise from the fact that the survey was not conducted in a part of the target population. Errors of observation are the differences between respondents' answers and true values. This distinction is in line with the commonly used dichotomy of representativeness errors versus measurement errors. There are three main types of

Figure 1. Eurostat's classification of sources of statistical errors for sample surveys



Source: Author according to Eurostat [2009].

non-observation error: coverage error, non-response error, and sampling error. Errors of observation include respondent error, interviewer error, instrument error, and error of data collection mode.

For a better understanding of contemporary approaches to survey error, Figure 1 depicts the Eurostat's concept [Eurostat 2009]. This concept differs from that of Groves by maintaining the basic distinction between sampling and non-sampling errors. Moreover, it contains a separate group of 'processing errors'. Different emphases are also placed on different types of error, with sampling errors occupying a privileged position. Individual types of measurement error are not elaborated separately in the basic scheme. However, in this case also the concept of statistical accuracy goes far beyond the original mathematical statistical concepts of sampling error.

The Eurostat concept of total error has specific practical applicability. It is not only described and explained, but, above all, Eurostat [2009] has produced a description of best practices for statistical bureaus and a handbook on the ways of reporting the quality of statistics delivered to the European Statistical System, including the indicators and level of detail that must be monitored in different types of research. Eurostat's approach to data quality monitoring has a strong quantitative leaning. However, the Eurostat system also foresees the application of nonprobability techniques for many surveys and accepts alternative evaluation methods and discussions of survey error.

What are the specifics of these various types of survey errors? Their description can be found in several conceptual or introductory works [e.g. Groves 2004; Lessler and Kalsbeek 1992; Biemer and Lyberg 2003; Groves et al. 2004], and during recent decades also a number of analytical studies on survey errors' impact on accuracy under different survey situations have been carried out by these and many other authors.

Probability sampling assumes non zero chances of selection for all individuals. Contrary to this, in the real world of survey research parts of studied populations usually remain out of reach of researchers. Coverage error reflects the fact that there is no chance of including some persons into the sample. The target population whose characteristics we infer in our conclusions is usually different from the 'survey population'. Our definition of the latter reflects our resignation to cover only certain groups that are with reasonable effort and cost available for interview. For example, people who live permanently in institutions or are homeless are typically excluded from general population surveys because such individuals are difficult to contact and interview.

Furthermore, the survey population usually differs from the 'sampling frame population'. It includes only those units that are available on the lists of the sampling frame or that can be reached by sampling techniques if there are no such lists. Often the sampling frame is not clearly defined. Differences arise not only from uncovered units of the target population but also from unwanted units in the sampling frame (e.g. in household research, wrong addresses or addresses of units other than households), duplicate units (e.g. several addresses for one household), clustered units (e.g. several households at one address), etc.[6] Moreover, many surveys combine two or more sampling frames by integrating different types of information, or apply complex sampling procedures. This results in overlapping frames where different units have different odds of being selected.

Thus, the coverage error reflects (1) the fact that some units of the target population are not included in the sampling frame population, and (2) differences in the statistic measured between the group of units included in the sampling frame and the group of those missing from it. This also means that if the values in group of persons missing in the sampling frame are not significantly different

[6] For some groups, we have difficulty in allocating a household to a specific residential unit. For example, some young people live simultaneously at both their parents' home and in their own home.

from the overall value of the population no error arises, even if there is a large share of such persons in the target population. Hence, coverage error is defined as a function of the difference between the target population value and the sampling frame population value.

A non-response error results from failure to collect data from all the units in the sample. This form of error may occur for a number of distinct reasons that may be summarised as follows. Not all persons selected are willing to answer questions in the questionnaire; not all respondents or households are reached by the interviewer during data collection; there are situations in which respondents are unable to participate (illness, language problems, etc.); and various organisational/administrative mistakes occur during data collection.

Random sampling usually makes it possible to calculate a survey's response rate. This is defined as the proportion of participating units in the eligible units sampled or, more specifically, the number of sufficiently completed interviews with respondent units as a proportion of the total number of eligible respondent units within the sampling frame. In this case, eligibility refers to a unit's membership in the sampling frame population. For instance, wrong addresses, units that are not sampled, and other units falling within the coverage error are excluded [Groves 2004; Lessler and Kalsbeek 1992; AAPOR 2009].

The response rate is often viewed as a key indicator of survey data quality. Therefore, it is sometimes subject to special efforts in order to make data more trustworthy. Nevertheless, data processing can rely on standard procedures for calculating the response rate that have been determined by the AAPOR and WAPOR associations[7] [AAPOR 2009] or Eurostat [2009]. Those standards are de facto binding for the member organisations of associations such as the ESOMAR or EFAMRO[8] [Smith 2002]. Procedures for calculating the response rate are also defined in numerous survey methodology works [Groves 2004; Lessler and Kalsbeek 1992].

However, the value of any non-response error can only be calculated theoretically. This is because the error is not directly related to missing observations, and instead, is co-determined by the level of differences between survey respondents and non-respondents. These differences may or may not be substantial. If the values observed in respondents were the same as those in non-respondents then a low response rate would not cause any error [Groves 2004; Groves and Couper 1998; Stoop and Louwen 2000]. At the same time, different sources of non-participation have different relationships with different variables, making non-response error specific for each statistic, rather than for the survey as a whole

[7] AAPOR is an acronym for the American Association for Public Opinion Research; WAPOR is an acronym for the World Association for Public Opinion Research.
[8] ESOMAR is the World Association for Opinion and Marketing Research Professionals; EFAMRO is an acronym for the European Federation of Associations of Market Research Organizations.

[Lessler and Kalsbeek 1992; Tourangeau, Rips and Rasinski 2000; Curtin, Presser and Singer 2000]. Analyses have also confirmed that the relationship between response rate and total error is not a direct one and may even be completely insignificant [Groves 2006; Keeter et al. 2000; Martin 2004; Brehm 1993]. Therefore, surveys with low response rates are not necessarily poor surveys. No generally applicable benchmark for an acceptable response rate can be defined in order to classify surveys. Similarly, any response rate increase that is achieved does not necessarily decrease the total error of the estimation.

Sampling error arises out of the heterogeneous character of the values of any statistics measured from any population. If a statistic is determined from a specific sample, its value is very likely to be different within a given range of values if many additional sample estimates were measured in an identical manner. Then the sampling error is a function of the difference in the distribution of the values within the sample and the distribution of the values within the sampling frame population. More specifically, sampling error stems from different values in different identically drawn samples taken from a given population. In the case of genuine probability samples[9] sampling error can be expressed numerically. Sampling error is the most frequent type of survey-based error examined and typically underlies statistical analyses of sample survey data.

It should be noted that in social analysis the uses of sampling error for statistical testing are quite specific. Notwithstanding the fact that many analysts make inappropriate use of statistical estimators, as Soukup and Rabušic [2007] highlight, the tests of statistical significance are usually based on hypothetical values of sampling error for a hypothetical survey based on simple random sampling. In reality, most survey datasets examined are not random samples and so the sampling error ranges assumed are incorrect. This statistical estimation problem is compounded by the fact that widely used statistical software assume by default that the data examined is a random sample, which is of course rarely the case.

In reality, most probability sampling surveys of general populations are based on 'complex samples', combining the procedures for stratified, cluster, and other sampling methods, which results in sampling errors of considerably different sizes. The 'design effect' is a ratio between the sampling variance of the statistic obtained by using a specific sampling procedure and the sampling variance that would be obtained for a given statistic by using simple random sampling. Effective sample size is defined as the size of a simple random sample that would produce the same sampling variance as the design currently applied.

Groves [2004] understands measurement errors as observational gaps between the ideal measurement and the response obtained. It may be added that these errors occur during data collection. In this respect, Groves distinguishes between four types of measurement error that are associated with inaccurate re-

[9] Probability samples are only those ones where sampling probabilities are known for all units in the sample.

cording of answers to questions from the research instrument. As with the representativeness errors or the missing observation errors discussed earlier, only a brief outline will be provided here.

Interviewer error corresponds to the effects of different interviewing practices on respondents' answers. Above all, the interviewer carries out multiple roles in the research process. He or she may participate in constructing the sampling frame, implementing the sampling procedure, and recruiting respondents, all of which may contribute to several types of errors (coverage, sampling, and non-response errors). Furthermore, the interviewer interacts with respondents, informing them of how they are expected to perform in the survey and helping them assume the appropriate role. Subsequently, he or she controls the interviewing process and records the answers. The mere presence of an interviewer constitutes an inevitable source of bias.

Respondent error arises out of the interviewee's inability to answer a question correctly, the interviewer's lack of effort to obtain a correct answer, or other psychological reasons. On the one hand, different kinds of respondents have different approaches to formulating answers. On the other hand, interviewing represents a complex cognitive process for respondents. In order to provide true answers, they must interpret and understand the question, find relevant information in their memory, assess the information, conduct estimations, and communicate the answer in a way that it is understood correctly.

Instrument error refers to the formulation of questions in standardised questionnaires as a source of bias and measurement variance. Here one must consider the wording of a question (the terms and formulations used), the structure of the questions and the measurement instrument as a whole (length of a question or set of questions, question types: open-ended/closed-ended, the number and order of response categories, the existence and formulation of 'don't know' categories, the flow of questions in terms of sensitive issues, etc.), and the context in which the question is applied (the position of a question within the questionnaire, other content of the questionnaire, etc.). This type of error is typically discussed within various extended concepts of measurement validity and reliability.

Mode error is associated with the characteristics of different modes of data collection. The basic types are as follows: the face-to-face interview between interviewer and respondent, the self-administered questionnaire, the telephone interview, and the Web-based interview. Many technological procedures that are currently implemented bring about a substantial change in the traditional ways of interviewing, sometimes combining their basic features. They include, for instance, computer-assisted personal interviewing (CAPI), computer-assisted self-interviewing (CASI), audio computer-assisted self-interviewing (ACASI), interactive voice response (IVR or T-ACASI), computer-assisted Web interviewing (CAWI), or computer-assisted telephone interviewing (CATI).

Apart from the communication channels and technologies used, the different methods also vary on the basis of interviewer participation and confidentiality. Thus, mode choice often substantially affects the occurrence of different types of errors as well as survey costs. It cannot be determined in advance which interviewing method is generally better. Different methods are suitable for different research situations. The interview topic and instrument complexity are also relevant. A vast analytical literature deals with the effects of mode choice on errors. The 'marginal mode effect' is determined by comparing two possible methods in a given research situation based on a given set of characteristics.

Processing errors often remain outside the interest of survey methodologists. They are related to technical operations of data entry, keying, coding, editing, and the conversion of a raw data file into the state needed for the intended data analysis. These processes are sometimes purely technical in nature, can be automated, and allow the establishment of systems of computer controls. On the other hand, they can also include highly complicated tasks in many different areas such as the imputation of missing data, weighting, the construction and transformation of indicators, the harmonisation of variables across nations and cultures, the integration of data from different sources, the solution to disclosure issues, or even the coding of some variables (e.g. the coding of occupations), etc. Some of these processes are based on quite sophisticated concepts and can be sources of substantial errors.

Quality standards

The above discussion of survey errors demonstrates that any in-depth evaluation of data quality requires an analytical study that is difficult, costly, and unfeasible under normal research circumstances and has to be based on detailed knowledge of the survey process. The goal of an in-depth quality evaluation thus exceeds the possibilities of individual research exercises and the reasonable size of a research report's methods section. Furthermore, often there is an information asymmetry between data collectors and analysts using their data. When data users are not directly involved in data collection, they have a limited ability to obtain all the necessary information about a survey and a limited ability to verify the information received.

As in other fields, the quality of the results in research is achieved by assuring the quality of the production process. As with many other products, the client of a survey agency has limited scope to exercise control over quality in any mass surveying process. Thus, as in other economic sectors, various standards help break down the information barrier between producers and users/customers.

While standards do not lend themselves to assuring the accuracy of a given survey's results, they affirm that the organisation that is trying to satisfy such standards has a structure capable of achieving a certain level of quality in the survey production process, abides by the rules for achieving and improving quality, and is subject to certain controls. In the following part of the article a brief overview of survey quality standards will be provided, for which purpose standards can be divided into two different types: (1) administratively imposed standards, and (2) professional associations' technical and ethical requirements for membership.

(a) Administrative standards

The aim of administrative standards is to regulate market conditions beyond legal rules in order to create a background (1) for clients to verify a product offer's compliance with national and international customs, so that they can assess the product's quality and compare offers, (2) for complaints about products and services delivered with reference to compliance with accepted rules, and (3) for the market and opinion polling research field's self-regulation.

Certification by the International Organization for Standardization (ISO) represents the most important system of internationally applicable standards. As a relatively new addition to the field of social research, ISO standard 20252:2006 for market, opinion, and social research has been published. This standard governs business relations and aims for the application of 'Total Quality Management' in social research. Within this framework, it foresees [Zahradníček 2006]:

- What the client is authorised to require of those conducting the research
- How to implement the research ethic into different stages of the survey research process and into the relations between survey agencies and their clients and survey agencies and survey respondents
- What any research brief must clearly fulfil
- What the client has to provide to those conducting the research so that the outcome of their cooperation is effective and efficient

Before 2006, the role of this framework was played by standards formulated by the ISO 9000 group, which generally focused on business quality management issues. Both ISO 20252:2006 and ISO 9000 are based on providing an independent assurance of an organisation's management structures in terms of its ability to fulfil quality requirements, apply regulatory procedures for this purpose, and ensure a continuous improvement process. The principles of quality management are reflected in the definition of the different dimensions of research quality and the formulation of binding procedures that are structured and phased. Company rules of conduct are adopted for the purposes of quality control in the research process and at individual management levels in order to achieve good quality at all times and facilitate continuous improvement.

Certification does not guarantee the accuracy of specific results or the high level[quality] of specific methodological procedures. However, it does guarantee that a research agency has an appropriate structure for the given type of research and is capable of achieving sufficient quality at every research stage.

(b) Standards of professional associations

The standards adopted by professional associations provide frameworks for the evaluation of data quality and comparison. In addition, professional standards define the rights of clients (as well as survey respondents), and provide guarantees for clients that member organisations will adhere to the application of specific standards. More encompassing and detailed rules with direct application to sample surveys have been created by a number of international social research organisations. Among them are Eurostat, the International Statistical Institute (ISI), the OECD, the UN, the IMF, the International Labor Organization (ILO), and many others.

Beyond the domain of official statistical services, the following two international professional associations in the fields of market, opinion and social research have also defined important standards. Both organisations enforce their standards in close mutual cooperation:

- ESOMAR (World Association for Opinion and Marketing Research Professionals)[10] is a worldwide association of scholarly organisations in market, opinion, and social research. Producers, distributors, and users of research results are among its members. It primarily focuses on the business sector. With more than 4600 members in more than 100 countries, it constitutes the most important association in the field.
- WAPOR (World Association for Public Opinion Research) brings together scholars in public opinion research. It works in approximately sixty countries.

Social research takes many forms. Surveys are conducted under various research circumstances. At the same time, contradictory requirements are sometimes placed upon the desired features of the resulting data. Therefore, no exhaustive and unequivocal set of criteria can be set. Two complementary approaches are available for overcoming this problem. On the one hand, minimal standards are defined. While no professionally conducted survey should ever break such standards, they often do not alone guarantee satisfactory quality. On the other hand, definitions of best practices are constructed. Such definitions are not binding but serve as useful guidelines and frames of reference.

In this respect, the ESOMAR has published an ICC/ESOMAR International Code on Market and Social Research [ESOMAR 2008], which is further elaborated and specified in a set of application notes [ESOMAR 2009]. The association continuously publishes and updates a large system of detailed guidelines for individual research domains and specific situations. The Code, including the application notes, is binding for ESOMAR members and has been recognised as an International Chamber of Commerce standard.

[10] The acronym used by this association is an abbreviation of its original name: the European Society for Opinion and Marketing Research. This name is no longer used because ESOMAR evolved from a primarily European into a global organisation.

WAPOR publishes various recommendations for research practice, and it also takes part in preparing and enforcing other organisations' standards. Together with ESOMAR it has developed guidelines for public opinion polls and publication of survey results [ESOMAR/WAPOR 2009] and closely cooperates with AAPOR (American Association for Public Opinion Research) in introducing its standards [AAPOR 1997, 2009] globally. This fact is crucial, because in this way the WAPOR transfers the relatively detailed and stringent US standards into international practice.

The Association of Market Research Agencies, SIMAR, enforces the standards of social research in the Czech Republic. SIMAR has published its own standards, which include a number of relatively specific methodological guidelines and minimum standards that refer to ESOMAR standards and ISO 20252:2006. SIMAR standards are binding for member organisations. Consequently, membership in the organisation represents a specific form of data quality assurance.

Conclusion

It should be recognised that the real circumstances of survey fieldwork often do not allow for the collection of data that would conform to the parameters required by the theoretical assumptions of many methods in mathematical statistics. Does this mean that we have to give up what is currently the most important source of data in empirical social research? Of course not; we have to make the best of the information that is achievable in the real world. Nevertheless, this fact means that the application of quantitative methods in the social sciences is not as simple and its results not as explicit as in luckier fields of research with a sufficient quality of data. That is why much more emphasis must be put on increasingly precise theoretical work and the construction of an additional knowledge base for the statistical analysis of social data. Social analysis that lacks solid grounding in knowledge of the studied problem is only an exercise in statistical methods and provides no reliable conclusions regarding social reality.

Undoubtedly these include also knowledge of the origins, methodological characteristics, and possible limitations of the data analysed. In this article, I demonstrated that such knowledge is impossible without taking a realistic view of the issue of data quality, which is much more complex than is assumed by those traditional approaches which consider only basic statistical principles. Unfortunately, in this regard, current practice in quantitative social research is not without problems. The following four principles, which might help to improve the situation, stem from the discussion of quality issues in this article:

First, it is important to promote and disseminate knowledge on the current concepts of survey data quality among researchers, followed by its implementation into the practice of social analysis. Current approaches have already become well established in the field of official statistics and have also penetrated the practice of data collection in other areas, but among academic researchers they are surprisingly uncommon.

Second, the establishment of data-sharing standards among the social science community, in order to ensure not only the public accessibility of data, but also the availability of appropriate survey documentation, represents a key goal. If there are no relevant barriers for their dissemination, such as legal or ethical barriers of copyright or personal data protection, the data should be made available together with appropriate documentation for purposes of secondary analysis and/or verification of research results.[11]

Third, it is crucial that submissions to scholarly publications give adequate attention to the origin and quality of the data analysed. Of course, it is not possible to require a complex evaluation of data quality in the limited space allocated to articles in scholarly journals, but at least basic information should be included. A comprehensive list of the elements proposed for the disclosure of the methods used in the survey is included in the AAPOR Best Practices [1997], and minimal standards can be obtained, for example, from ESOMAR, WAPOR or SIMAR codes [ESOMAR 2009; ESOMAR/WAPOR 2009; SIMAR 1996–2008].

Fourth, promoting the establishment of survey quality norms and standards is an important objective. Such formal and informal principles are important tools for reducing the asymmetry in knowledge regarding the data collection process between data producers and data users.

To sum up, survey quality represents one of the central pillars of empirical social research. Failure to ensure survey quality owing to a lack of knowledge undermines the whole social scientific enterprise. It is hoped that this article contributes to highlighting this fundamentally important issue and encourages researchers (a) to pay greater attention to how their datasets are (and should be) produced, and (b) to employ such knowledge in their data analysis, causal inferences, and published work.

JINDŘICH KREJČÍ *is the head of the Sociological Data Archive (SDA) of the Institute of Sociology, Academy of Sciences of the Czech Republic. His research focuses on survey methodology, data quality, and the study of political attitudes. He is the author of the book* Kvalita sociálněvědních výběrových šetření v České republice *(The Quality of Social Science Sample Surveys in the Czech Republic, Prague 2008) and the author of several articles in scientific journals and books. At the Faculty of Social Sciences of Charles University he teaches courses in methods of social data analyses and data management. He has also participated in surveys undertaken for the International Social Survey Programme (ISSP) and the European Social Survey (ESS).*

[11] These principles are supported in the guidelines produced by the OECD and EU, e.g., *The OECD Principles and Guidelines for Access to Research Data from Public Funding* [OECD 2007].

References

AAPOR. 1997. *Best Practices for Survey and Public Opinion Research*. Deerfield: The American Association for Public Opinion Rsearch. Retrieved 5 August 2010 (http://www.aapor.org/Best_Practices/1480.htm).

AAPOR. 2009. *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. Deerfield: The American Association for Public Opinion Research. Retrieved 5 August 2010 (http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions&Template=/CM/ContentDisplay.cfm&ContentID=1819).

Andersen, R., J. Kasper and M. Frankel (eds.). 1979. *Total Survey Error*. San Francisco: Jossey-Bass.

Babbie, Earl. 2004. *The Practice of Social Research*. 10th ed. Belmont: Thomson Learning.

Bell, John F. 1991 'Big Is Not Necessarily Beautiful in Survey Design: Measurement Error and the APU Science Survey.' *The Statistician* 40 (3: Special Issue: Survey Design, Methodology and Analysis 2): 291–300.

Biemer, Paul P. and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken: John Wiley & Sons.

Blalock, Hubert M., Jr. 1989. 'The Real and Unrealised Contributions of Quantitative Sociology.' *American Sociological Review* 54: 447–460.

Brehm, John. 1993. *The Phantom Respondents: Opinions Surveys and Political Representation.* Ann Arbor: University of Michigan Press.

Colledge, Michael and Mary March. 1997. 'Quality Policies, Standards, Guidelines, and Recommended Practices at National Statistical Agencies.' Pp. 501–522 in *Survey Measurement and Process Quality*, edited by L. E. Lyberg, P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz and D. Trewin. New York: John Wiley.

Curtin, R., S. Presser and E. Singer. 2000. 'The Effects of Response Rate Changes on the Index of Consumer Sentiment.' *Public Opinion Quarterly* 64 (4): 413–428.

Dalenius, Tore. 1977. 'Bibliography of Nonsampling Errors in Surveys.' *International Statistical Review* 45: 71–81, 181–197.

Dalenius, Tore. 1981. 'The Survey Statistician's Responsibility for both Sampling and Measurement Errors.' Pp. 17–29 in *Current Topics in Survey Sampling*, edited by D. Krewski, R. Platek and J. N. Rao. New York: Academic Press.

Denis, Daniel J. 2003. 'Alternatives to Null Hypothesis Significance Testing.' *Theory & Science* 14 (1). Retrieved 17 August 2010 (http://theoryandscience.icaap.org/content/vol4.1/02_denis.html).

ESOMAR. 2008. *ICC/ESOMAR International Code of Marketing and Social Research Practice*. ESOMAR World Research Codes and Guidelines. Amsterdam: ESOMAR. Retrieved 5 August 2010 (http://www.esomar.org/uploads/pdf/professional-standards/ICCESOMAR_Code_English_.pdf).

ESOMAR. 2009. *Notes on How to Apply the ICC/ESOMAR International Code of Marketing and Social Research Practice*. ESOMAR World Research Codes and Guidelines. Amsterdam: ESOMAR. Retrieved 5 August 2010 (http://www.esomar.org/uploads/professional_standards/guidelines/ESOMAR_Codes&Guidelines_NotesHowToApplyCode.pdf).

ESOMAR/WAPOR. 2009. *ESOMAR/WAPOR Guide to Opinion Polls and Published Surveys*. ESOMAR World Research Codes and Guidelines. Amsterdam: ESOMAR. Retrieved 5 August 2010 (http://www.esomar.org/uploads/professional_standards/guidelines/WAPOR-ESOMAR_Guidelines.pdf).

Eurostat. 2002. *Quality in the European Statistical System – The Way Forward*. Luxembourg: Office for Official Publications of the European Communities. Retrieved 5 August

2010 (http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/KS-45-02-814/EN/KS-45-02-814-EN.PDF).
Eurostat. 2003. *Methodological Documents – Definition of Quality in Statistics*. Working Group on Assessment of Quality in Statistics, Luxembourg, 2–3 October 2003. Luxembourg: Eurostat. Retrieved 5 August 2010 (http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf).
Eurostat. 2009. *ESS Handbook for Quality Reports*. Luxembourg: Office for Official Publications of the European Communities. Retrieved 5 August 2010 (http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-016/EN/KS-RA-08-016-EN.PDF).
Gigerenzer, G., S. Krauss and O. Vitouch. 2004. 'The Null Ritual: What You always Wanted to Know about Significance Testing but Were Afraid to Ask.' Pp. 391–408 in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, edited by D. Kaplan. Thousand Oaks, CA: Sage.
Gill, J. 1999. 'The Insignificance of Null Hypothesis Significance Testing.' *Political Research Quarterly* 52 (3): 647–674.
Groves, Robert M. [1989] 2004. *Survey Errors and Survey Costs.* 2nd ed. Hoboken: John Wiley & Sons.
Groves, Robert M. 2006. 'Nonresponse Rates and Nonresponse Bias in Household Surveys.' *Public Opinion Quarterly* 70 (5): 646–675.
Groves, R. M. and M. P. Couper. 1998. *Non-Response in Household Interview Surveys.* New York: Wiley.
Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer and Roger Tourangeau. 2004. *Survey Methodology*. Hoboken: John Wiley & Sons.
Hansen, M. H., W. N. Hurwitz and L. Pritzker. 1967. 'Standardization of Procedures for the Evaluation of Data: Measurement Errors and Statistical Standards in the Bureau of the Census.' *Bulletin of the International Statistical Institute* (Proceedings of the 36th Session): 49–66.
ISO. 1994. *ISO 8402:1994. Quality – Vocabulary*. Geneva: International Organization for Standardization.
Juran, J. M. and Frank M. Gryna. 1993. *Quality Planning and Analysis: From Product Development Through Use*. 3rd ed. New York: McGraw-Hill.
Kaase, Max (ed.). 1999. *Qualitätskrieterien der Umfrageforschung – Quality Criteria for Survey Research*. Berlin: Akademie Verlag GmbH.
Keeter, S., C. Miller, A. Kohut, R. M. Groves and S. Presser. 2000. 'Consequences of Reducing Non-response in a National Telephone Survey.' *Public Opinion Quarterly* 64 (2): 125–148.
Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley & Sons.
Laliberté, Lucie, Werner Grünewald and Laurent Probst. 2004. *Data Quality: A Comparison of IMF's Data Quality Assessment Framework (DQAF) and Eurostat's Quality Definition*. IMF Reference Material. Retrieved 5 August 2010 (http://dsbb.imf.org/vgn/images/pdfs/dqaf_eurostat.pdf).
Lessler, Judith T. and William D. Kalsbeek. 1992. *Nonsampling Errors in Surveys*. New York: John Wiley & Sons.
Lynn, Peter. 2001. *A Quality Framework for Longitudinal Studies*. Reference Document of the ESRC National Strategy Committee for Longitudinal Studies. Colchester: ESRC. Retrieved 5 August 2010 (http://natstrat.essex.ac.uk/pages/documentation/reference/framework.doc).
Martin, E. 2004. 'Presidential Address: Unfinished Business.' *Public Opinion Quarterly* 68 (3): 439–450.

Merton, Robert K. and Paul F. Lazarsfeld (eds.). 1950. *Continuities in Social Research: Studies in the Scope and Method of 'The American Soldier'*. Glencoe: Free Press.
Morganstein, David and David A. Marker. 1997. 'Continuous Quality Improvement in Statistical Agencies.' Pp. 475–500 in *Survey Measurement and Process Quality*, edited by L. E. Lyberg, P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz and D. Trewin. New York: John Wiley.
Noelle, Elisabeth. 1963. *Umfragen in der Massengesellschaft*. *Einführung in die Methoden der Demoskopie.* Reinbek bei Hamburg: Rowohlt Tagensbuch Verlag.
OECD. 2002. *Quality Framework for OECD Statistics*. Paris: OECD. Retrieved 5 August 2010 (http://www.oecd.org/document/43/0,3343,en_2649_33715_21571947_1_1_1_1,00.html).
OECD. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: Organisation for Economic Co-Operation and Development. Retrieved 5 August 2010 (http://www.oecd.org/dataoecd/9/61/38500813.pdf).
SIMAR. 1996–2008. *Kvalitativní standardy* (Quality standards). On-line version. Retrieved 5 August 2010 (http://www.simar.cz/standardy-kvality/kvalitativni-standardy/index.php).
Smith, Tom W. 2002. 'Developing Nonresponse Standards.' Pp. 27–40 in *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J. A. Little. New York: John Wiley & Sons.
Soukup, Petr and Ladislav Rabušic. 2007. 'Několik poznámek k jedné obsesi českých sociálních věd – statistické významnosti.' (Some Notes on the Obsession of the Czech Social Sciences with Statistical Significance) *Sociologický časopis/Czech Sociological Review* 43 (2): 379–395.
Statistics Canada. 2002. *Statistics Canada's Quality Assurance Framework*. Ottawa: Statistics Canada. Retrieved 5 August 2010 (http://www.statcan.gc.ca/pub/12-586-x/12-586-x2002001-eng.pdf).
Stoop I. A. L and F. Louwen. 2000. 'Do Respondents Differ?' Paper presented at the Eleventh International Workshop on Survey Nonresponse, 27–29 September, Budapest, Hungary.
Tourangeau, Roger, Lance J. Rips and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
Zahradníček, Stanislav. 2006. *Komentář k ČSN 20252:2006 'Výzkum trhu a veřejného mínění a sociální výzkum.'* (A Commentary on the ČSN 20252:2006 'Market, Public Opinion and Social Research') Prague: Český normalizační institut.