Survey Measurement and Process Quality

Edited by

Lars Lyberg
Paul Biemer
Martin Collins
Edith de Leeuw
Cathryn Dippo
Norbert Schwarz
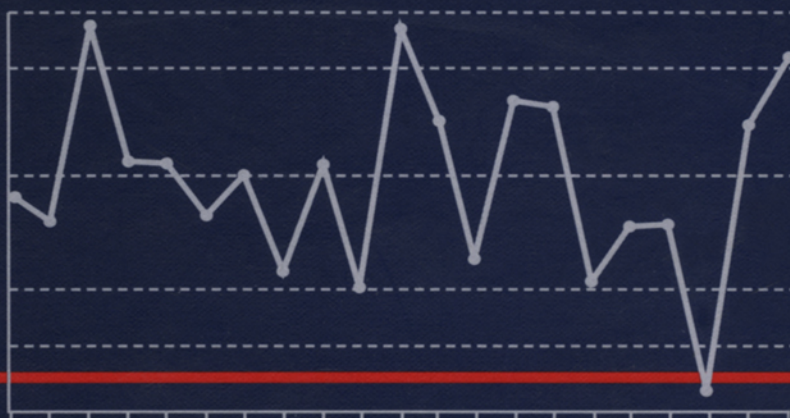**Dennis Trewin**

Survey Measurement
and Process Quality

Survey Measurement and Process Quality

Edited by

LARS LYBERG
PAUL BIEMER
MARTIN COLLINS
EDITH DE LEEUW
CATHRYN DIPPO
NORBERT SCHWARZ
DENNIS TREWIN

Contents

SECTION E. ERROR EFFECTS ON ESTIMATION, ANALYSES, AND INTERPRETATION

Contributors

Reginald P. Baker, National Opinion Research Center, Chicago, Illinois, U.S.A.

Alison K. Baldwin, National Opinion Research Center, Chicago, Illinois, U.S.A.

Francesca Bassi, University of Padua, Padua, Italy

Mary K. Batcher, Internal Revenue Service, Washington, DC, U.S.A.

Jelke G. Bethlehem, Statistics Netherlands, Voorburg, The Netherlands

Paul P. Biemer, Research Triangle Institute, Research Triangle Park, North Carolina, U.S.A.

Steven Blixt, MBNA-America, Wilmington, Delaware, U.S.A.

Bill Blyth, Taylor Nelson AGB, London, United Kingdom

Pamela Campanelli, Social Community Planning Research, London, United Kingdom

Noel Chavez, University of Illinois, Chicago, Illinois, U.S.A.

Michael Colledge, Australian Bureau of Statistics, Belconnen, Australia

Martin Collins, City University Business School, London, United Kingdom

Frederick Conrad, Bureau of Labor Statistics, Washington, DC, U.S.A.

Mick P. Couper, University of Michigan, Ann Arbor, Michigan, U.S.A.

Edith de Leeuw, Vrije Universiteit, Amsterdam, The Netherlands

Don A. Dillman, Washington State University, Pullman, Washington, U.S.A.

Cathryn S. Dippo, Bureau of Labor Statistics, Washington, DC, U.S.A.

Jennifer Dykema, University of Wisconsin, Madison, Wisconsin, U.S.A.

James L. Esposito, Bureau of Labor Statistics, Washington, DC, U.S.A.

Luigi Fabbris, University of Padua, Padua, Italy

Leandre R. Fabrigar,   Queen's University at Kingston, Ontario, Canada

Wayne Fuller,   Iowa State University, Ames, Iowa, U.S.A.

Leopold Granquist,   Statistics Sweden, Stockholm, Sweden

Bill Gross,   Australian Bureau of Statistics, Belconnen, Australia

Patricia M. Guenther,   Department of Agriculture, Washington, DC, U.S.A.

Sue Ellen Hansen,   University of Michigan, Ann Arbor, Michigan, U.S.A.

K. Piyasena Hapuarachchi,   Statistics Canada, Ottawa, Canada

Anthony Heath,   Nuffield College, Oxford, United Kingdom

John Horm,   National Center for Health Statistics, Hyattsville, Maryland, U.S.A.

Joop Hox,   University of Amsterdam, Amsterdam, The Netherlands

Cleo R. Jenkins,   Bureau of the Census, Washington, DC, U.S.A.

Jared B. Jobe,   National Center for Health Statistics, Hyattsville, Maryland, U.S.A.

Timothy Johnson,   University of Illinois, Chicago, Illinois, U.S.A.

Daniel Kasprzyk,   Department of Education, Washington, DC, U.S.A.

John G. Kovar,   Statistics Canada, Ottawa, Canada

Jon A. Krosnick,   Ohio State University, Columbus, Ohio, U.S.A.

Jouni Kuha,   Nuffield College, Oxford, United Kingdom

Loretta Lacey,   University of Illinois, Chicago, Illinois, U.S.A.

Rolf Langeheine,   University of Kiel, Kiel, Germany

James M. Lepkowski,   University of Michigan, Ann Arbor, Michigan, U.S.A.

Susan Linacre,   Australian Bureau of Statistics, Belconnen, Australia

Lars E. Lyberg,   Statistics Sweden, Stockholm, Sweden

Mary March,   Statistics Canada, Ottawa, Canada

David A. Marker,   Westat Inc., Rockville, Maryland, U.S.A.

Jean S. Martin,   Office for National Statistics, London, United Kingdom

Nick Moon,   NOP Social and Political, London, United Kingdom

David R. Morganstein,   Westat Inc., Rockville, Maryland, U.S.A.

Colm O'Muircheartaigh,   London School of Economics and Political Science, London, United Kingdom

William L. Nicholls II,   Bureau of the Census, Washington, DC, U.S.A.

Diane O'Rourke,   University of Illinois, Urbana, Illinois, U.S.A.

Sarah M. Nusser,   Iowa State University, Ames, Iowa, U.S.A.

William F. Pratt,   National Center for Health Statistics, Hyattsville, Maryland, U.S.A.

Thomas W. Pullum,   University of Texas, Austin, Texas, U.S.A.

J. N. K. Rao,   Carleton University, Ottawa, Canada

Kenneth A. Rasinski,   National Opinion Research Center, Chicago, Illinois, U.S.A.

Jennifer M. Rothgeb,   Bureau of the Census, Washington, DC, U.S.A.

Sally Sadosky,   University of Michigan, Ann Arbor, Michigan, U.S.A.

Fritz J. Scheuren,   George Washington University, Washington, DC, U.S.A.

Norbert Schwarz,   University of Michigan, Ann Arbor, Michigan, U.S.A.

Jacqueline Scott,   Queen's College, Cambridge, United Kingdom

Randy R. Sitter,   Carleton University, Ottawa, Canada

Chris Skinner,   University of Southampton, Southampton, United Kingdom

Tessa Staples,   Office for National Statistics, London, United Kingdom

S. Lynne Stokes,   University of Texas, Austin, Texas, U.S.A.

Seymour Sudman,   University of Illinois, Urbana, Illinois, U.S.A.

Katarina Thomson,   Social Community Planning Research, London, United Kingdom

Roger Tourangeau,   National Opinion Research Center, Chicago, Illinois, U.S.A.

Dennis Trewin,   Australian Bureau of Statistics, Belconnen, Australia

Frank van de Pol,   Statistics Netherlands, Voorburg, The Netherlands

Richard Warnecke,   University of Illinois, Chicago, Illinois, U.S.A.

Adam Wronski,   Statistics Canada, Ottawa, Canada

Michaela Wänke,   University of Heidelberg, Heidelberg, Germany

Preface

Survey quality is directly related to survey errors. Survey errors can be decomposed in two broad categories: sampling and nonsampling errors. Comprehensive theories exist for the treatment of sampling errors. As for nonsampling errors, no such theory exists. Indeed, there has been extensive interest in nonsampling error research over the last 50 years which has resulted in an abundance of literature describing the treatment of various specific error sources and attempts at an integrated treatment or simultaneous modeling of several specific error sources. In some ways this research has been very successful and has given rise to efficient methods of dealing with some error sources. On the other hand, many types of errors have exhibited a resistance to the suggested solutions. One reason, of course, is that errors are generated by a vast number of sources, which is then further complicated by the great complexity of some survey designs.

Another reason is that nonsampling errors usually are not additive. Reducing the level of one type of error might very well increase some other type of error. Indeed, some survey quality goals are conflicting, which introduces an irrational element into the decision making. For instance, attempts at reducing non-response rates by intense follow-ups might sabotage timeliness; wealth of detail regarding survey data might violate confidentiality safeguards; and reducing processing errors might call for questionnaires that are less user-friendly. Admittedly, some error sources can be dealt with, but others are so elusive that they defy expression, not to mention estimation.

Left uncontrolled, nonsampling errors can render the resulting survey data useless for many important survey objectives. Post-survey quality assessment measures such as reliability, validity, and bias estimates are very important indicators of data accuracy, but, except for repeated surveys, may be of little value for improving the survey data. Rather, interest must shift from post-survey quality evaluation and possible correction to controlling the survey processes such as questionnaire construction, interviewing and other data collection activities, coding, data capture, editing, and analysis. Process quality generates product quality.

Many survey organizations throughout the world are now working with the concepts of Total Quality Management (TQM) in the context of survey design, data collection, and data processing. Methods for monitoring and ensuring process quality such as process control, quality teams, customer focus, decisions based on scientific methods, and so on which have been developed in industrial settings are being successfully applied in survey work. Very little of this work is reported in the literature yet the potential of these methods is enormous.

Given the importance of the topic of survey measurement and process quality, the Survey Research Methods Section (SRM) of the American Statistical Association (ASA) in 1992 determined that survey measurement and process quality should be the topic of an SRM-sponsored conference and edited monograph and approached Lars Lyberg to develop the idea. It was decided that the conference should seek the participation of researchers worldwide and that it should take place in a European country to ensure a wider dissemination of research findings and exploit the great interest emerging in countries outside the U.S. which, geographically, has been the locus of interest in this topic. Also, the SRM emphasized the need for interdisciplinary and cross-cultural research.

By early 1993, an organizing/editing committee was formed consisting of: Lars Lyberg (Chair), Paul P. Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. Lee Decker was enlisted to be in charge of the conference logistics and most administrative issues related to the conference. Patricia Dean was enlisted as consulting editor for the monograph. Lilli Japec was enlisted to be in charge of contributed papers and other planning tasks.

The committee contacted numerous research organizations for financial contributions. The committee also developed the monograph outline and began to identify and contact researchers throughout the world as potential authors. Abstracts were requested and 133 abstracts were received from researchers interested in writing for the monograph. From these, the committee selected 34 which would be the chapters in this monograph and developed the conference program. Steve Quigley orchestrated John Wiley & Sons' role in publishing this monograph.

Five professional organizations were asked to sponsor the conference: the American Statistical Association (ASA), the International Association of Survey Statisticians (IASS), the Market Research Society (MRS), the Royal Statistical Society (RSS), and the World Association for Public Opinion Research (WAPOR). All five organizations enthusiastically agreed and also contributed funds to support the project. In addition, the following research organizations contributed funds:

Australian Bureau of Statistics
Central Statistical Office, Dublin
Economic & Social Research Council, London

International Labour Office, Geneva
National Opinion Resarch Center, Chicago
NSS, The Hague
National Science Foundation, Washington, DC
Office for National Statistics, London
Research Triangle Institute, Research Triangle Park
SPSS, London
Statistics Denmark
Statistics Finland
Statistics Netherlands
Statistics New Zealand
Statistics Sweden
Survey Research Center, Institute for Social Research, University of Michigan
U.S. Bureau of Labor Statistics
U.S. Bureau of the Census
U.S. Department of Agriculture/NASS
U.S. National Center for Health Statistics
Westat, Inc., Rockville
ZUMA, Mannheim

Without the financial support of these organizations, the conference and edited monograph would not have been possible.

The International Conference on Survey Measurement and Process Quality was held April 1–4, 1995 in Bristol, United Kingdom. It drew 291 attendees from 27 countries. The program consisted of the 34 invited papers chosen for the present monograph and 113 contributed papers. Authors of contributed papers were offered the opportunity to submit their papers to a proceedings volume published by the American Statistical Association. The volume contains 60 of the contributed papers. Additionally, two short courses were offered, one on multilevel analysis for survey research and one on TQM in statistical organizations. The organizing committee was also very pleased by the fact that the Sir Ronald Fisher Memorial Committee of Great Britain chose our conference for its XIXth Fisher Memorial Lecture. This lecture is included in the proceedings volume.

In designing this book, the aim has been to discuss the most important issues in the field of survey measurement and process quality, attempting whenever possible to integrate various perspectives. Thus, each chapter has undergone extensive editing, review, and revision. The book is organized into five sections. The section titles and their editors are:

Section A: Questionnaire Design (Norbert Schwarz)
Section B: Data Collection (Edith de Leeuw and Martin Collins)
Section C: Post Survey Processing and Operations (Lars Lyberg)
Section D: Quality Assessment and Control (Cathryn Dippo)

Section E: Error Effects on Estimation, Analyses, and Interpretation (Paul Biemer and Dennis Trewin).

The diversity of orientations of the authors for the monograph made it impossible to impose a unified terminology and set of notation across all chapters. Except for Section E, the statistical level of the monograph is quite accessible by graduate students in sociology, psychology, communication, education, or marketing research. Section E, however, requires a fairly thorough foundation in survey sampling and mathematical statistics.

Although the present book can serve as a course text, its primary audience is researchers having some prior knowledge in survey research. Since it contains a number of review articles on survey measurement and process quality in several disciplines, it will be useful to researchers actively involved in this field who want a discussion from different theoretical perspectives. The book will also be useful to methodologists who want to learn more about improving the quality of surveys through better design, data collection, and analytical techniques and by focusing on processes. The book reflects current knowledge in 1995, to the best of our editorial judgment. As a group, we hope that its publication will stimulate future research in this exciting field.

Most section editors had responsibilities as the secondary editor for one other section as well. The authors of the chapters, in addition to their extensive writing and revising activities, were also involved in the review of other monograph chapters. They were encouraged to seek outside reviews for their chapters on their own. Thus, the monograph reflects the efforts and contributions of scores of writers, editors, and reviewers. The committee would like to sincerely thank Patricia Dean who performed the final editing of all manuscripts. She served as consulting editor and her efforts went far beyond regular language editing in providing authors with suggestions regarding style and organization of all chapters.

We are again grateful to Lee Decker of the ASA, who handled all the logistical details associated with the conference and the proceedings volume with great care and efficiency. Lilli Japec deserves great appreciation for all the activities she performed so ably for the conference and monograph.

Sincere thanks go to Joke Hoogenboom and Truus Kantebeen at the SRM Documentation Centre at the Erasmus University in Rotterdam, who compiled a booklet "SMPQ, A Selected Bibliography" that was available at the conference. They also did literature searches for all sections of the monograph.

Thanks are due to Barbara Bailar, Seymour Sudman, and Judith Lessler, who while serving in various ASA positions promoted the idea of an international conference on this topic to be held outside the U.S., which is the usual locus for these activities. We are appreciative of the efforts of Dan Kasprzyk, Jun Liu, Joop Hox, John Bosley, Fred Conrad, Sylvia Fisher, Roberta Sangster, Linda Stinson, and Clyde Tucker who assisted the committee in the review of a number of chapters. Sincere thanks go to Eva von Brömssen who compiled the initial list of index entries. Our employing organizations also deserve great

appreciation for supporting our activities in conducting the conference and assembling the monograph: Statistics Sweden (Lyberg); University of Michigan (Schwarz); Vrije Universiteit, Amsterdam (de Leeuw); City University (Collins); U.S. Bureau of Labor Statistics (Dippo); Research Triangle Institute (Biemer); and Statistics New Zealand and Australian Bureau of Statistics (Trewin).

<div align="right">

LARS LYBERG
PAUL BIEMER
MARTIN COLLINS
EDITH DE LEEUW
CATHRYN DIPPO
NORBERT SCHWARZ
DENNIS TREWIN

</div>

Survey Measurement
and Process Quality

INTRODUCTION

Measurement Error in Surveys: A Historical Perspective

Colm O'Muircheartaigh
London School of Economics and Political Science

1 ORIGINS OF SURVEY ASSESSMENT

In considering the different contexts in which early users of surveys operated, and the different perspectives they brought to their operations, it is difficult to find common criteria against which to measure the success of their endeavors. The history of surveys (in their modern sense) goes back only 100 years, but from the outset there was a great diversity in the settings, topics, philosophies, and executing agencies involved. In the initial stages there was no particular distinction drawn between the issues of survey design and execution and the issues of error in surveys.

   The concept of quality, and indeed the concept of error, can only be defined satisfactorily in the same context as that in which the work is conducted. To the extent that the context varies, and the objectives vary, the meaning of error will also vary. I propose that as a definition of error we adopt the following: *work purporting to do what it does not do*. Rather than specify an arbitrary (pseudo-objective) criterion, this redefines the problem in terms of the aims and frame of reference of the researcher. It immediately removes the need to consider *true value* concepts in any absolute sense, and forces a consideration of the needs for which the data are being collected. Broadly speaking, every survey operation has an objective, an outcome, and a description of that outcome. Errors (quality failures) will be found in the mismatches among these elements.

   There are three distinct strands in the historical development of survey research: governmental/official statistics; academic/social research; and

commercial/advertising/market research. Each of these brought with it to the survey its own intellectual baggage, its own disciplinary perspective, and its own criteria for evaluating success and failure.

The International Statistical Institute (ISI) was the locus of debate for official statisticians at the end of the 19th century when Kiaer, director of the Norwegian Bureau of Statistics, presented a report of his experience with "representative investigations" and advocated further investigation of the field. In this context the evaluation of surveys was largely statistical and the survey was seen as a substitute for complete enumeration of the population. Bowley—the first professor of statistics at the London School of Economics and Political Science—through his work on sampling (1906 and 1926) and on measurement (1915) was one of the principal figures in the development of the scientific sample survey. This became and has remained the dominant methodology in the collection of data for government, and the government sample survey agency became an important purveyor of data both to politicians and to statesmen. Symptomatic of their genesis, these agencies tended to be located in the national statistical office, and their professional staff tended to be trained in mathematics or in statistics. Here the concept of error became synonymous with the variance of the estimator (essentially the variance of the sampling distribution following Neyman's influential paper in 1934 (Neyman, 1934)). This equivalence of quality and variance and its measurement by repeated sampling, with some acknowledgment of bias, was confirmed by the work of Mahalanobis in India, reported in the mid-1940s (see Mahalanobis, 1944, 1946), and in particular by his design of interpenetrating samples for the estimation of fluctuations or variability introduced by fieldworkers and others. The influence of statisticians on the conceptualization of error and its measurement has continued in this tradition, and can be found in all the classic texts of survey sampling (Yates, 1949; Cochran, 1953; Hansen et al., 1953; Kish, 1965). In this tradition the term "error" has more than one meaning (see, for example, Groves (1989)) but it is used loosely to describe any source of variation in the results or output or estimates from a survey.

While recognizing the powerful position occupied by the scientific sample survey in social research, it is worth noting that Kiaer's proposal to the ISI in 1895 was not universally welcomed, and would almost certainly have been rejected had it not been for the support of the monographers whose work consisted of the detailed examination of one or a small number of cases (what might today be called the *case study* approach).

The involvement of the monographers in the debate at the ISI is interesting particularly because it provides a link to the second major strand in the development of surveys. This was the *Social Policy* and *Social Research* movements, whose beginnings are perhaps best represented by Booth's study, from 1889 to 1903, of poverty in London, and the Hull House papers in the U.S.A. in 1892. Though not in any way a formal or organized movement, there were certain commonalities of approach and objectives across a wide range of activities. The goal of this movement was social reform, and the mechanism

was community description. Here the success or failure of the activity was the effect the findings had on decision makers and politicians.

The principal influences on this group were the social reform movement and the emerging sociology discipline. Some of the pioneers of sample surveys spanned both official statistics and the social survey; in particular, Bowley (who made a substantial contribution to the development of survey sampling) produced a seminal work on social measurement in 1915 which helped define the parameters of data quality and error for factual or behavioral information. Bogardus (1925), Thurstone (1928), and Likert (1932) provided scientific approaches to the measurement of attitudes. In this field the disciplinary orientation was that of sociology and social psychology, with some influence from social statistics and psychometrics. Likert, who was subsequently the founding director of the Institute for Social Research at the University of Michigan in 1946, reflected the same practical orientation as the early pioneers in the Social Research movement in his later work on organizations (though with extensive use of attitude measurement).

The third strand arose from the expansion of means of communication and growth in the marketplace. From modest beginnings in the 1890s (Gale and others), there was a steady increase in the extent of advertising and a development and formalization of its companion, market research. The emphasis was on commercial information, originally in the hands of producers of goods and services and collected and evaluated informally (Parlin, 1915); market research, however, developed gradually into a specialized activity in its own right.

Here the effect of psychologists was particularly strong. The work of Link and others in the Psychological Corporation was influential in providing an apparently scientific basis for measurement in the market research area. For those psychologists, experimental psychology took precedence over social psychology. The terminology and the approach were redolent of science and technology. The term "error" was not used explicitly; rather there was a description of *reliability* and *validity* of instruments. This contrasts particularly with the "error" orientation of the statisticians.

Thus the field of survey research as it became established in the 1940s and 1950s involved three different sectors—government, the academic community, and business; it had three different disciplinary bases—statistics, sociology and experimental psychology; and it had developed different frameworks and terminologies in each of these areas.

2 FRAMEWORK

In general in describing data quality or errors in surveys, models concentrate on the survey operation itself, in particular on the data collection operation. The models may be either mathematical (presenting departures from the ideal as disturbance terms in an algebraic equation) or schematic (conceptual models

describing the operational components of the data collection process). The conceptual models focus on the interview as the core of the process. Building on the work of Hyman (1954), Kahn and Cannell (1957), Scheuch (1967) and others, Sudman and Bradburn present one of the more useful of these in their book on response effects in surveys (Sudman and Bradburn, 1974). This (schematic) model presents the relationship among the interviewer, the respondents and the task in determining the outcome of the survey interview. The elaborated model identifies the potential contribution of a number of the key elements in each of these to the overall quality of the survey response.

- The *interviewer*, as the agent of the researcher, is seen to carry the lion's share of responsibility for the outcome of the data collection process. Sudman and Bradburn distinguish three elements: the formal constraints placed on the interviewer; the actual behavior of the interviewer; and the extra-role characteristics of the interviewer.
- The *respondent* has not generally been examined as a source of error (apart from a general complaint about poor performance of his/her task). Survey research has tended to take the respondent for granted, though many of the early writers referred to the need to motivate the respondent. The overall approach has, however, been to consider the respondent an obstacle to be overcome rather than an active participant in the process.
- In general, models of response errors in surveys focus on the *task*, which is constrained and structured to accomplish the research goals—in particular to provide the data necessary for analysis. The task includes the length and location of the interview, the question wording, questionnaire construction, the types of data sought, and their implications in terms of memory and social desirability.
- The *responses* are the outcome of the data collection exercise, and the raw material for survey analysis. Most survey analyses treat these as free from error; the statistical approach to measurement error considers the response to be a combination of the *true value* of the data for the individual plus a disturbance described as a *response deviation* or *response effect*.

It is clear that any model of the survey process, and therefore any general model of survey error, will have to include these elements. It is not, however, sufficient to consider these elements, as they do not take into account the context of a survey nor can they distinguish among different survey objectives. To compare the different approaches to survey research described in Section 1, however, it is necessary to provide an over-arching framework that encompasses the concerns of all three major sectors.

One possible framework draws on some ideas presented by Kish in his book on statistical design for research (Kish, 1987). Kish suggests that there are three issues in relation to which a researcher needs to locate a research design; I propose that a similar typology could be used to classify the dimensions that

would encompass most sources of error. Each of these "dimensions" is itself multi-dimensional; they are *representation, randomization,* and *realism.*

As survey research deals with applied social science, our understanding of measurement in surveys must also be grounded in actual measures on population elements. Social theory does not have this requirement, nor indeed does statistical theory. At this empirical level, however, the strength and even the direction of relationships between variables are always conditional on the elements, and thus it is critical that any conclusions from a survey should be based on a suitable set of elements from the population, and that comparisons between subclasses of elements should be based on comparable subsets of elements. *Representation* involves issues such as the use of probability sampling, stratification, the avoidance of selection bias, and a consideration of non-response. In general we do not believe that any finding in social science will apply uniformly to all situations, in all circumstances, for all elements of the population. Indeed a good deal of social science is dedicated to understanding the ways in which differences occur across subgroups of populations or between populations. *Representation* reflects this set of concerns with regard to the elements included in the investigation. In particular it refers to the extent to which the target population is adequately mirrored in the sample of elements. In a perfectly specified model, there would be no need to be concerned about which elements from the population appeared in the sample. In the absence of complete and perfect specification of a model (with all variables with potential to influence the variables or relationship under consideration being included), the notion of representation specifically covers the appropriate representation of domains (or subclasses), the avoidance of selection bias, and the minimization of differential nonresponse. The term representative sampling has a chequered history in statistics (see, for instance, Kruskal and Mosteller, 1980; O'Muircheartaigh and Soon, 1981). It carries with it an aura of general (possibly) unjustified respectability; it can be taken to mean the absence of selective forces (that could lead to selection biases); its original connotation was that of a miniature or mirror of the population; it has sometimes been seen as a typical or ideal case; it can imply adequate coverage of the population (cf. stratification); its highest manifestation is in probability sampling, which is the approach in academic and (most) governmental research.

Randomization (and its converse in this context, *control*) covers issues of experimentation and control of confounding variables. Though surveys rarely involve the use of formal experiments for substantive purposes, the identification of sources of measurement error (distortion in the data) and the estimation of the magnitudes of these "errors" frequently do. Randomization is used to avoid, or reduce the probability of, spurious correlations or mis-identification of effects. (It may be worth pointing out that randomization (or at least random selection) is also used to achieve representation in probability sampling.)

Realism arises as an issue in this context in two ways. *Realism in variables* concerns the extent to which the measured or manifest variables relate to the constructs they are meant to describe; *realism in environment* concerns the degree

to which the setting of the data collection or experiment is similar to the real-life context with which the researcher is concerned. The survey context may be contrasted with observational studies in which both the variables and the environment are closer to the reality we would like to measure. These dimensions are related to the ideas of *internal validity* and *external validity* used by Campbell and Stanley (1963) and others in describing the evaluation of social research. The validity of a comparison within the context of a particular survey is the realm of internal validity; the extent to which an internally valid conclusion can be generalized outside that particular context is the realm of external validity.

In the following sections the different components of the response process are presented. Each of them concentrates on a different element of the basic model. Section 3 presents the perspective of official (government) statistics and concentrates on the *responses*; this tradition is still followed by the *hard science* school of survey research. Section 4 considers the elements of the *task*. Section 5 takes as its focus first the *interviewer*, then the *respondent* and the interrelationship between them; in Sections 4 and 5 most of the contributions to progress have been made by either the psychologists involved in market and opinion research, or by the sociologists and social psychologists involved in social and policy research. In Section 6 some recent developments are used to illustrate how measurement error in surveys is being reconsidered. Section 7 presents some tentative conclusions.

3 THE EFFECT OF THE SAMPLING PERSPECTIVE— STATISTICAL ANALYSIS OF THE RESPONSES: OFFICIAL STATISTICS AND SURVEY STATISTICS

The sample survey was seen by Kiaer (1897) and its other originators in government work as an alternative to complete enumeration necessitated by the demand for more detail and more careful measurement. In 1897 Kiaer wrote "In order to arrive at a deeper understanding of the social phenomena . . . it is necessary to . . . formulate a whole series of special questions . . . prohibitive to conduct a complete survey of the population of a country, indeed even one for all the inhabitants of a large town" (p. 38). It was the necessity to *sample* that brought about the difference between the survey and the usual government enquiry, and it was the errors that might contaminate the results because of this that became the focus of concern for the first generation of statisticians and others involved with government surveys. Kiaer suggested *replication*— simply constructing a set of comparable subsamples (in essence repeating the sampling operation)—as the means of evaluating the survey results (p. 51); this was, as far as I know, the first *total variance model*.

This approach was taken on board by Bowley and other statisticians and culminated in the classic 1934 paper by Neyman to the Royal Statistical Society "On the Two Different Aspects of the Representative Method" which

crystallized the ideas in his concept of the *sampling distribution*—the set of all possible outcomes of the sample design and sampling operation. The quality of a sample design, and thus a sample survey, could be encapsulated in the *sampling error* of the estimates; it is worth noting that though the general term "error" was used, this was a measure purely of variance or variability, and not necessarily a measure of error in any general sense. In particular, *bias* (or systematic error) was not a primary consideration, except as a technical issue in the choice among statistical estimators.

The Kiaer–Bowley–Neyman approach produced the sequence of texts on sampling which have defined the social survey field for statisticians ever since. The sequence of classic sampling texts began with Yates (1949, prepared at the request of the United Nations Sub-Commission on Statistical Sampling for a manual to assist in the execution of the projected 1950 World Census of Population), and Deming (1950), followed by Hansen *et al.*, (1953), Cochran (1953), and Sukhatme (1953), and concluded with Kish (1965). With these texts survey statisticians defined their field as that of measuring the effect of sample design on the imprecision of survey estimates. Where other considerations were included they tended to be relegated to a subsidiary role, or confined to chapters towards the end of the book. The texts vary a good deal in terms of the relative weight given to mathematical statistics; the most successful as a textbook, however, Cochran (2nd edition 1963, 3rd edition 1977) was the most mathematical and the least influenced by nonsampling and nonstatistical concerns.

A second strand was present in the work of Mahalanobis in India. He, like Kiaer, advocated the use of replication, using what he called *interpenetrating samples*, to estimate the precision of estimates derived from a survey. He defined these as "independent replicated networks of sampling units." He was, moreover, the first statistician to emphasize the *human agency* in surveys (1946, p. 329); he classified errors as those of sampling, recording, and physical fluctuations (instability). To estimate variance, he advocated that different replicates should be dealt with by "different parties of field investigators" so that human error as well as sampling errors would be included in the estimates of precision; he also carried out tests of significance among replicates. Mahalanobis may also be credited with perceiving that an additional advantage of partial investigations (using his interpenetrating samples) was that they facilitated the estimation of error, something previously not a part of the reports of government agencies. Indeed one of his early evaluations, based on sampling by the Indian Civil Service between 1923 and 1925, showed a bias in the estimation of crop yields (1946, p. 337).

Replication remains the primary instrument of statisticians when dealing with error. The traditional division between *variance*—the variability across replications, however defined—and *bias*—systematic deviation from some correct or true value—still informs the statistician's approach to error. Replication (or Mahalanobis's interpenetration) is the method of producing *measurability* in the sense of being able to estimate the variability of an estimate from within the process itself. In sampling, this was brought about by selecting a

number of sampling units independently and using the differences among them as a guide to the inherent stability or instability of the overall estimates. For simple response variance, the statisticians simply repeated the observations on a subset of respondents and compared the outcomes; this is usually called a reinterview program; this gives replication in the sense of repetition. In the context of interviewer effect, the replication is within the survey operation and is brought about by constructing comparable subsets of cases and comparing them. To measure interviewer effect, respondents are allocated at random to different interviewers and the responses obtained are compared. Statistical theory tells us how much variability we could expect among these interviewer workloads if there is no systematic interviewer effect on the responses. To the extent that the variation is larger than that predicted by the null model, we attribute the effect to the interviewers.

The early 1960s saw the next step forward in statisticians' consideration of survey error. Hansen *et al.* (1961) in a seminal paper presented what became known as the "U.S. Census Bureau" model of survey error. They defined the *essential survey conditions* as the stable characteristics of the survey process and the survey organization carrying it out; variance was defined relative to those essential survey conditions. The observation is seen as being composed of two parts, its *true value*, and a deviation from that value—the *response deviation*. Though Hansen and his colleagues were well aware that the notion of a "true value" is problematic, they proposed it as a useful basis for the definition and then estimation of error. Their model is essentially a variance-covariance model and permits considerable generalization (see, for example, Fellegi, 1964, 1974) and has been extremely influential among survey statisticians. In particular it allows the incorporation of the effects of interviewers and supervisors, and the possibility of correlated errors within households.

About the same time, Kish (1962) presented findings using an alternative technical approach using analysis of variance (ANOVA) models in dealing with interviewer error; this was the approach favored by Yates, among others, and derived from the experimental design perspective of agricultural statisticians. Again the statistician simplifies reality so that it can be accommodated within the structure of his/her models; the effect of the interviewer is seen as an additive effect to the response of each respondent interviewed by that interviewer. The approach is easily generalizable to the effects of other agents in the survey (coders, supervisors, editors, etc.) (see, for instance, Hartley and Rao, 1978); one drawback is that the ANOVA models do not easily lend themselves to the analysis of categorical data.

These two approaches have in common the objective of estimating the variance of the mean of a single variable (or proportion). The focus of a survey is seen as the estimation of some important descriptive feature of the population. Thus the total variance is seen as the sum of the various sources of error affecting the estimate. Starting with the variance of the sample mean of a variable measured without error and based on a simple random sample (SRS), a set of additional components may easily be added to the variance, each representing