SAGE researchmethods

Single- and Mixed-Mode Survey Data Collection

Foundation Entries

SAGE Research Methods Foundations

By: Vera Toepoel, Edith de Leeuw & Joop Hox
Published:2020
Length: 10,000 Words
DOI: http://dx.doi.org/9781529748987
Methods: Single- and Mixed-Mode Survey Data Collection
Online ISBN: 9781529748987
Disciplines: Anthropology, Business and Management, Criminology and Criminal Justice, Communication and Media Studies, Counseling and Psychotherapy, Economics, Education, Geography, Health, History, Marketing, Nursing, Political Science and International Relations, Psychology, Social Policy and Public Policy, Social Work, Sociology, Science, Technology, Computer Science, Engineering, Mathematics, Medicine
Access Date: November 17, 2020
Publishing Company: SAGE Publications Ltd
City: London

Abstract

The main data collection modes for surveys are the face-to-face interview, telephone interview, paper mail survey, and online survey. This entry deals with both single- and mixed-mode survey data collection. Survey modes are associated with different types of errors. These types of errors can be based in the representation side, that is, whether or not the sample represents the population, or be related to measurement issues. This entry discusses the total survey error paradigm and how this affects the different modes of data collection. Logistics associated with single- and mixed-mode surveys are discussed, as well as the use of auxiliary data and methods to analyze (mixed-mode) surveys. The entry ends with a view toward the future of single- and mixed-mode data collection, including mixed-application integrated research that combines survey research with other types of data.

Introduction

This section introduces surveys and data collection mode, and how they relate to total survey error (TSE).

Surveys and Data Collection

In his seminal booklet, Fritz Scheuren (2004) described a survey as a method of gathering (quantitative) information from a sample of individuals. Robert Groves and colleagues (2009) define a survey as a method for gathering information from (a sample of) entities for the purpose of constructing quantitative descriptors of the attributes of the larger population of which the entities are members. Many disciplines use surveys as a data collection method to answer research questions. Surveys may attempt to measure everyone in a population or just a sample. In some cases when the population is very small, no sample is drawn and the total population is investigated; an example is a health survey among patients with a very rare disease. Most often, for logistics and budgetary reasons, and to avoid the risk of oversurveying a population, a subset (sample) of the population is investigated. To be able to draw valid statistical inferences about the population of interest, the sample should be a probability-based sample (Lohr, 2008). A questionnaire is then administered to the sample, either by an interviewer or through a self-completion method.

The main data collection modes for surveys are the face-to-face interview, telephone interview, paper mail survey, and online survey. Each mode has its own advantages and disadvantages. In addition, a combination of data collection modes can be used. This entry deals with both single- and mixed-mode survey data collection. The choice for the optimal survey mode is often made on the basis of the intuition of the researcher. However, it is important to make an informed decision and to know the different types of errors that come with a choice for a particular survey mode. These types of errors can be based in the representation side, that is, whether or not the sample represents the population, or be related to measurement issues.

This entry first discusses the TSE paradigm and how this affects the different modes of data collection (face-to-face, telephone, paper mail, and online). Next, mixed-mode surveys, that use a combination of modes, are taken into account. Further, logistics associated with single- and mixed-mode surveys are discussed, as well as the use of auxiliary data and methods to analyze (mixed-mode) surveys. The entry ends with a view toward the future of single- and mixed-mode data collection, including mixed-application research that combines survey research with other types of data.

Total Survey Error

The TSE framework is the central organizing structure in the field of survey methodology. When a researcher is planning a survey, there are practical constraints such as costs, ethics, and time. When conducting a survey, several errors may occur, resulting in a deviation of the value estimated based on the survey and the actual value in the population. TSE emphasizes the trade-offs that must be made in trying to minimize the potential survey errors within the constraints of the available resources such as budget and time (Groves & Lyberg, 2010; Lyberg & Weisberg, 2018). A promising value of the TSE framework lies at the design stage of a survey since it attempts to balance costs and various errors.

TSE components link steps in the measurement and representational inference process. It distinguishes six types of errors: three on the representation side (coverage error, sampling error, and nonresponse error) and three on the measurement side (specification error, measurement error, and processing error). In the first step of a survey, the target population is defined, and the sampling frame is described. Coverage error refers to a mismatch between the intended target population and the sampling frame. For instance, the sampling frame may not cover all individuals in the population (undercoverage), and this difference between the target population and the sampling frame may cause coverage error. The next step in the survey process, from sampling frame to actual sample, may cause sampling error. The fact that not all individuals in the population are part of the actual group (sample) studied may cause an error in estimates: the sampling error. Sampling error can only be estimated when a probability-based sample is used; that is, when each individual has a known and nonzero probability to be sampled. Once an individual is selected in the sample, he or she is asked to participate in the study. Since not all invited individuals participate, for various reasons (e.g., could not be contacted, refused, ill), all summarized in the term nonresponse, this may result in nonresponse error. Nonresponse error occurs when there is nonresponse *and* the respondents and nonrespondents differ on variables of interest in the study.

Representation error is concerned with the selection of respondents and their agreement to cooperate. Measurement-related error is related to the question–answer process itself. Specification error occurs when the concept measured through the survey question differs from the intended, theoretical concept; this refers to the validity of the measurement. Measurement error itself arises whether there is a discrepancy between the measurement of a construct and the actual (true) value. Finally, processing error may arise when the answer provided by the respondent is edited by the researcher or data collection agency, for example, when a response provided on paper is coded and entered in a digital file. Combining these six different types of

error gives the overall error, the TSE. Researchers should be aware that the survey statistic in their data set is affected by TSE.

TSE and Data Collection Mode

Each survey mode has its own associated survey errors. Face-to-face surveys have the advantage of an interviewer. An interviewer can persuade people to participate in the study and can help and motivate them. Although an interviewer may help clarify certain questions, this clarification can also systematically bias results. Also, interviewers may induce socially desirable responding, particularly in sensitive questions. Therefore, face-to-face surveys score better on coverage and nonresponse error but may do worse on measurement error. To compensate for socially desirable responding in face-to-face interviews, a self-administered component can be added to the survey for sensitive questions (mixed-mode approach). Using a computer-assisted personal interview can help limiting processing errors.

For telephone surveys, it is important to include both landline and mobile phone numbers. When both are included, telephone surveys perform well on coverage error. The response rate, and hence response error, depends on the country, but response can be quite low. Telephone surveys resemble face-to-face surveys since in both modes there is an interviewer present. However, a telephone interviewer is more limited than a face-to-face interviewer in time and means persuading respondents to cooperate. Furthermore, telephone surveys are more prone to measurement error due to the limited channels of communication. In a face-to-face survey, both aural and visual communication can be used: The respondent sees and hears the interviewer, and the interviewer may show visual material such as show cards with lists of answer categories. In a telephone survey, only aural communications is possible.

Paper mail survey coverage error depends on the sampling frame that is being used. In countries with good quality and accessible registers (e.g., northern European countries), or with other systems such as postal delivery files (e.g., United States), coverage error can be quite low. Paper mail surveys have reasonably good response rates (somewhere between face-to-face and online). A disadvantage of paper mail surveys is that there is no control or help for respondents. In addition, paper mail surveys have to be processed, scanned, and coded with the danger of added processing error. Well-written and tested questions and layout are paramount to limit measurement error, and questionnaires with a lot of branching (skip questions) are difficult from a respondent viewpoint because he or she has to understand the branching without the help of an interviewer or an automated program. However, as all self-administered modes, paper mail questionnaires perform better when sensitive questions are asked and have less social desirable answers than interviews. Paper mail surveys also have lower costs than interviewer administered modes but are typically more expensive than online surveys.

Online surveys tend to be the least expensive mode since variable costs (costs per complete) are minimal. As fixed costs (e.g., programming and testing) remain the same, independent of the number of completed questionnaires, the costs savings are larger for very large samples. The online survey has to be programmed, and there is much variation in layout and user-friendliness in online survey software. How survey software

deals with different devices—and small screen sizes in mobile phones in particular—differs greatly between survey vendors. Online surveys suffer from poor response rates, and a lack of a suitable sampling frame (there is no list with e-mail addresses of the general population in most countries) makes them prone to either nonprobability sampling or probability-based off-line sampling. Measurement and processing error tend to be low in well-designed and tested online surveys since the online survey program takes care of processing such as skip questions (although this depends on the program used).

There are no systematic differences when it comes to sampling error and the four different modes. To profit from the low selection error in interviewer administered modes, and the low measurement error in the self-administered mode, often a combination of modes is used. These studies are called mixed-mode studies. Any modes can be mixed in different combinations.

Mixed-Mode Design

This section introduces mixed-mode surveys, how to perform mixed-mode surveys and consequences of mixed-mode surveys. In addition, it provides examples of how to design questions for mixed-mode surveys.

Definition of Mixed Mode

From a TSE point, the main reasons to implement a mixed-mode survey are to reduce coverage and nonresponse error, and potentially improve data quality, all at affordable costs. In the first articles on mixed-mode design, the definition of mixed mode was broad and in addition to surveys that use multiple modes of data collection included surveys that combined a specific data collection method with different modes of contacting, recruiting, or reminding respondents. An example of the latter is a paper letter to announce the visit of an interviewer, explaining the utility of the survey and emphasizing the saliency and legitimacy. However, in the 2010s, such a broad definition would include almost all academic and official surveys. For instance, to combat low response rates, paper advance letters are typically used in telephone and face-to-face surveys, and to be able to use probability sampling in online studies, respondents for probability-based online panels are contacted and recruited through face-to-face or telephone interviews. Therefore, in the following discussion, the restricted definition of mixed mode is used, and designs are described for surveys that combine multiple modes of data collection (for a detailed taxonomy, including different modes for contact, reminders, and recruitment, see de Leeuw, 2005).

Using more than one data collection mode in one survey can be traced back to the 1960s. For many years, it has been standard good practice to offer a self-administered questionnaire for sensitive topics within a face-to-face interview. This specific *concurrent* mixed-mode design, whereby two data collection modes are used at the same time, has as a goal the reduction of measurement error due to social desirability. An early application of a *sequential* mixed-mode design to reduce nonresponse is a paper mail survey with a telephone interview as follow-up.

Sequential Versus Concurrent Design

In cross-sectional studies, a sequential mixed-mode design is mainly used to achieve a high response rate at affordable cost. Usually, the survey starts with the least expensive method, and increasingly more expensive methods are used for the nonresponse follow-up. A prime example is the American Community Survey that starts with online/mail survey, follows up with a telephone survey of nonrespondents, followed by a face-to-face interview of a subsample of the remaining nonrespondents. Sequential mixed-mode studies are also used in a longitudinal setting. However, in a longitudinal study, one usually starts with the more expensive method (e.g., face-to-face interview) to maximize coverage and minimize nonresponse and achieve a baseline study. The next waves are conducted with a less expensive mode to minimize overall costs. A special case of this longitudinal sequential approach is used for establishing a probability-based online panel. Lacking good sampling frames of e-mail addresses of the general population, a well-defined sampling frame is used (e.g., household addresses, telephone numbers), and potential panel members are recruited through traditional interview surveys. Examples are the Dutch LISS-panel, the German GESIS-panel, and the French ELIPSS.

To reduce noncoverage error and to include the non-Internet population, probability-based online panels use several approaches. The LISS-panel offered a simple PC and high-speed Internet to those without, ELIPSS provided all panel members (both those online and off-line) with a standard tablet, and the GESIS-panel offers a postal mail survey to panel members without Internet access.

The latter is a form of concurrent mixed-mode (online-mail) design. Another form of a concurrent mixed-mode design is offering potential respondents a choice between modes, for example, by sending a paper letter with an URL for an online survey and the offer to send a paper questionnaire when preferred. This approach may reduce coverage error as it is sent to both those online and off-line. Furthermore, the expectation was that this would increase response rates; the rationale is that different persons have different mode preferences and offering a choice is more respondent friendly and might boost the response. However, empirical evidence and meta-analyses show that offering a choice between web and paper mail results in lower response rates than a single mode mail survey (Medway & Fulton, 2012). Offering a choice may seem respondent friendly, but in fact, it places an extra burden on the potential respondent, who now has to make two decisions: Do I want to participate and in which mode do I want to participate? Confronted with this extra burden, the easiest way out is to procrastinate and do nothing at all. Therefore, more recent methods use a “nudge to the web” strategy that uses mail contacts to request response to an online survey and withholding other survey modes until later to contact nonrespondents (e.g., Dillman, 2017). However, when one really knows the preference of the respondent, for instance, from earlier contacts or segmentation analysis, tailoring to this preference and immediately offering the preferred mode is effective, as the GESIS-panel shows.

A natural form of concurrent mixed-mode design occurs in cross-national surveys. Countries differ in data collection traditions, infra structures (e.g., Internet penetration), population-specific needs, and field work requirements. A good example is the International Social Survey Program (ISSP), which started as a self-

administered paper questionnaire. When more countries entered the ISSP, the self-administered mode was unsuitable for low-literacy countries, and face-to-face interviews were offered in those countries.

As mentioned earlier, a special type of concurrent mixed-mode design is used to reduce measurement error in surveys on sensitive topics. In this design, all respondents get the same mix and different modes (interview and self-administered) are used for different parts of the questionnaire. This has the advantage that interviewers can motivate respondents and assist with complex question but switch to a more private self-administered mode for sensitive questions and so may reduce social desirability. Examples are handing over a paper questionnaire and envelope, or handing over the laptop in computer-assisted interviews.

Consequences of Mixed-Mode Surveys

Sequential mixed-mode strategies have been effective to improve response rates in cross-sectional studies designs, and there are indications that mixed-mode designs also work in longitudinal studies. Furthermore, there is evidence that mixed-mode surveys can reduce coverage error and improve representativeness (see de Leeuw, 2018).

Cost considerations are an important factor in the choice for mixed-mode design; however, much depends on the initial single-mode data collection. Using a mixed-mode strategy decreases fieldwork costs when compared to an expensive single-mode face-to-face design but not compared to an inexpensive web or paper-mail survey. When choosing a mixed-mode approach, researchers always should carefully balance survey errors (coverage, nonresponse, and measurement) and survey costs (Tourangeau, 2017).

Mixing modes has clear advantages to reduce coverage and nonresponse error at reasonable costs. When a survey includes sensitive topics, a special mixed-mode design is used in which all interviewees are offered a self-administered questionnaire for the sensitive part to reduce social desirability. But, in all other cases when different data collection modes are used for different groups of respondents, there is a danger of selective measurement error. For example, when online nonrespondents are followed up with interviews, usually younger respondents will more often respond online and elderly respondents will more often respond to interviews. Furthermore when sensitive questions on smoking behavior are asked and a (negative) correlation is found between age and smoking, then one does not know whether this is caused by age (younger people smoke more) or by mode (online more honest answers were given). This danger is larger when more disparate modes (e.g., online and interview) are mixed than when more comparable modes are combined (e.g., paper mail and online), and most measurement differences have been found between interview surveys on the one hand and self-administered modes on the other hand (see De Leeuw & Hox, 2011). Mode measurement effects can be enlarged by differential question format traditions used for certain modes such as offering do-not-know options or using grid questions; therefore, researchers should always design and implement a mixed-mode survey carefully.

Designing Equivalent Questionnaires

Originally, there were two different schools in survey design for mixed-mode studies. The first emphasized optimization of each mode separately to minimize measurement error within each mode even if that meant that different question formats were used in each mode. This approach advocated that when an *overall estimate* is wanted, measurement errors add up, and optimizing each mode separately leads to the best overall estimate. The second approach emphasizes equivalence and aims to minimize differences across modes. Modes should be comparable, and this is extremely important when data from a mixed-mode survey are used to compare different subgroups, countries, or institutions.

The second approach is often characterized as the unified or uni-mode approach, whereby the same question wording and structure is used across modes. Especially for questions that have been proved to be susceptible for question format changes, even within a single mode, such as attitudes, careful questionnaire construction across modes is essential.

In modern practice, these two schools are not as far apart as originally thought. Today, the emphasis is on designing equivalent questions and questionnaires to maximize data quality within a mode and minimize mode measurement effect. Equivalent does not necessarily mean the same word-by-word text. Just as a literal translation from one language to the next may not make sense, the same wording which is fine in one mode may seem stilted or strange in another. Therefore, whenever questionnaire designers have good reasons to deviate from using the same wording or format across modes, they should be allowed to do so. Sometimes differences in rules of communication and information transmission between modes force designers to deviate. A simple and straightforward example is the use of instructions or explanations for the respondent; in order to make sense these must be adapted to the mode. In an online interview, the text “click next to continue” or use of “arrow buttons” to go back and forward are necessary. In interviews explanatory or transitional texts need to be spoken out aloud, such as, “I will now read out a list of … . ” Also, the flow of the question–answer sequence and the rapport with the respondent need different requirements in interviewer-administered versus self-completion modes. If an interviewer has to read out the same instructions over and over again when asking a set of questions, this may seriously hamper the flow of the interview. Therefore, interviewer instructions in questionnaires often state, “repeat if necessary,” and a well-trained interviewer knows whether this is necessary or not, while the instruction in a self-administered questionnaire should always be directly accessible for the respondent.

It should be noted that equivalence does *not* mean scaling down to the lowest common denominator. The build-in intelligence of computerized questionnaires should be used, even if a paper questionnaire without these aids is also part of the mix. To increase the overall quality of the data and reduce the burden for the respondent and interviewer, automatic routings, fills, and checks should be used whenever possible.

Two Examples of Equivalent Questionnaire Design

Designing and implementing equivalent questionnaires demand a thorough knowledge of the individual

survey modes and a creative mind to combine the best of the different modes. This subsection provides two examples of how the need for equivalence led to a new and inventive question format. The first example concerns matrix questions in a combined online and interview survey, the second example is on the use of do-not-know options.

In visual modes, like an online survey, a series of questions with the same response categories is often presented in a matrix format, whereby the rows are the questions and the columns the common set of response options. All questions are presented simultaneously and visually together. In contrast, in face-to-face and telephone interviews questions are asked sequentially, one at a time. Although respondents are asked the same questions and are offered the same response options across modes, the questions are not equivalent. The respondents are offered different stimuli online and in the interview mode, which in this case may give rise to differential context effects across modes caused by adjacent questions.

An alternative to presenting a series of related questions in a matrix is the auto-advance question, also known as carrousel question, or horizontally scrolling matrix format. In this format, questions are presented one-by-one on the screen, mimicking the sequential interview process. The response options are shown in a horizontal line of boxes under the question. When the respondent has clicked on an answer, the next question flies in using auto-advance procedures; there is no need to click on a “next” button, thereby keeping respondent burden light. A visual navigation bar provides respondents with an overview of the number of questions answered and enables them to go back to a previous question. The auto-advance format is also helpful in the so-called mixed-device surveys, in which an online survey may be answered on a variety of devices such as PC, laptop, tablet, and mobile phone.

The second example addresses do-not-know options in interviews and self-administered surveys. In both telephone and face-to-face interviews, a do-not-know option is usually not explicitly offered to the respondent, but it is accepted whenever a respondent spontaneously answers “do not know.” Interviewers are trained to follow up a spontaneous do-not-know answer with a friendly probe to reduce uninformative answers. This format with an including probe is often programmed in as standard option for computer-assisted interviewing. However, in online surveys, web designers are inclined to either explicitly offer a do-not-know option or to omit the do-not-know answers and make the question mandatory. In a mandatory format, respondents have to provide an answer, and if they try to skip a question, they receive an—usually not so friendly—error message. Respondents will clearly *not* perceive these web formats in the same way as they perceive a friendly interviewer probe after a spontaneous do-not-know answer. Friendly interviewer probes are accepted by respondents, while error messages and a mandatory format online in which clicking on an answer category is obligatory may result in irritation and more guessing. In fact, these different formats result in different missingness patterns, and there are more do-not-know answers in Internet surveys with voluntary question format than in face-to-face surveys with friendly probes.

However, not probing in the interview mode is no option either because this will result in high item nonresponse in both interview and online modes. One solution is using the intelligence of a computerized interview environment, and in an online survey mimic an interviewer situation. In a respondent-friendly

approach, error messages can be formulated in such a way that they emulate interviewer probes or feedback. By analyzing interviewer instructions and interviewer text and using this in formulating online follow-up probes, it is possible to reduce the amount of item nonresponse to the same degree as in an interview with probing and keep respondent evaluations at a high positive level.

Inventive and thorough questionnaire design may reduce mode measurement effects, still some mode inherent effects, like privacy and social desirability, may remain. Thus, it is necessary to investigate the potential of mode effects and when necessary correct for them. This is discussed later in this entry.

Advantages and Disadvantages of Survey Modes in Logistics

In this section, the advantages and disadvantages of logistics are discussed for single- and mixed-mode surveys

Logistics for Single-Mode Surveys

Face-to-face surveys have the highest potential regarding types of questions asked and complexity of questionnaires. To realize this potential, one needs both well-trained interviewers and well-tested questionnaires. In addition, a highly qualified field staff is necessary to make sure that all logistics are taken care of. A face-to-face survey is very costly and time-consuming, especially if the country is large and sparsely populated. Cluster sampling may be needed, and if the sample dispersion is very high, telephone surveys may be preferred. Face-to-face surveys are only worth the money and time in situations in which accuracy is paramount, for example, for statistical offices.

Telephone surveys have less potential regarding types of questions than face-to-face interviews, as no visual communication is possible. But interviewers are available to help and guide the respondent and complex questionnaires may be used. However, fewer questions can be asked since telephone surveys should be around 15–30 min and no longer due to respondent fatigue. Due to unlisted numbers and mobile phones, coverage may be suboptimal. However, if good lists are available, telephone surveys are, from a sampling point of view, comparable to face-to-face interviewing. Nonresponse in telephone surveys is extremely high, resulting in relatively high costs per complete. Quality control in telephone surveys is good as interviewers can be closely monitored and immediate feedback is possible. Many interviews can be completed in a relatively short time with a smaller number of interviewers than in face-to-face surveys.

Mail surveys lack the flexibility and interviewer support of interview surveys, which limits the complexity of the questionnaire used. However, visual stimuli, such as pictures or graphics, can be applied and examples or show material may be included. Mail surveys are less intrusive than interviews: Respondents may answer at their own time and place. In addition, there is no interviewer present who may inhibit free answers to more

sensitive topics. Lists with addresses of the target population are typically available, creating the possibility for address-based sampling. Mail surveys have a longer turn around than telephone surveys, but face-to-face interviewing usually takes even longer. Mail surveys are far less costly than both face-to-face and telephone interview surveys since there is no need of interviewers or interviewer training.

For online surveys, much is dependent on the Internet penetration rate of the population. Internet access varies strongly between countries and within countries and large coverage errors may arise, as those who do not have access to the web are typically very different in terms of attitudes and behaviors than those who have Internet access. Lists with e-mail addresses of the general population are typically not available, although specific populations such as students and employees often have an e-mail address provided by the institution. In Internet surveys, complex questionnaires and visual stimuli can be easily applied, but questionnaires have to be very short. In an era of mobile surveys, the attention span of respondents is typically fewer than 10 min. With a range of different devices being used (e.g., PC, laptop, tablet, smartphone), pretesting on different devices, with different browses and screen sizes, is important. Like mail surveys, Internet surveys are less intrusive and can be taken at any time and place. Mobile phones make it possible to complete surveys on the move but still a majority answers at home or at work. Large numbers of completed questionnaires can be collected in a very short time and at low costs, making the online survey the preferred mode for most survey practitioners.

Logistics for Mixed-Mode Surveys

The reason for a mixed-mode strategy is typically to reduce coverage and nonresponse error, as well as measurement error. Conducting a mixed-mode survey can either reduce costs or increase costs, depending on the original single-mode survey and also on the sequence of the modes in the mix. No matter which combinations of modes are used, mixed-mode surveys always result in more complicated logistics.

Mixed-mode surveys add to the operational complexity because different recruitment strategies are necessary, for example, researchers need both home addresses and telephone numbers if they want to combine face-to-face and telephone surveys. Each mode has to be designed and tested, and it is wise for researchers to utilize integrated software that can deal with more than a single mode. In practice, a mixed-mode design often imposes a penalty during data preparation because it requires reformatting and joining disparate data files. Also, the analysis is more complicated as differential selection and measurement effects have to be taken into account.

To cope with the operational complexity, expertise is needed on all of the modes that are being used. This can be obtained with in-house expertise or by working with subcontractors. For each mode, expertise is needed on questionnaire development, questionnaire testing, fieldwork implementation, sampling issues, data management, and analysis issues. Often, experts on one mode work in a different department than the experts in the other mode, so careful communication is required. Questionnaire developers, programmers, statisticians, and data analysists need to work together to come up with plans for sampling, questionnaire development, fieldwork, data entry and coding, data merging, adjustment, and and analysis.

A total quality approach is related to total survey design, in which documentation is written down in field reports and disseminated. These field reports contain information on the survey process and quality, the methodologies that are being used, software, adjustment weights, and paradata (data that are being collected in addition to the survey data, such as keystroke files; see De Leeuw, Hox & Dillman, 2008).

Auxiliary Data

In this section, the need for auxiliary data in addition to the regular survey data is explained.

Nonresponse and Need for Auxiliary Data

To investigate and possibly to correct for survey errors, additional (auxiliary) data are needed. A prime example is nonresponse studies. Just reporting a response rate for a survey is not enough; nonresponse error should also be investigated. Nonresponse error occurs only when there are differences between respondents and nonrespondents on the variables of interest. To investigate nonresponse error, researchers need data about the respondents and the nonrespondents to be able to compare these two groups. Often demographic variables are used, such as gender, age, education, and urbanicity. When there are differences between respondents and nonrespondents on these variables, then this can be corrected for by using weights. Another example is the use of representativity or R-indicators together with the response rate. Here, auxiliary data on both respondents and nonrespondents are used to create a representativity indicator, which indicates how well the sample represents the population. The choice of the variables from the auxiliary data set is very important, and when an important variable that is related to the nonresponse mechanism is not included, weighting and R-indicators have limited value.

Auxiliary data are also necessary to estimate measurement errors. Validation studies use data from a benchmark source and compare survey results with these benchmark data. Many special studies into social desirability compared data from a survey with data from a second source, which are thought to be superior. For example, students' grades from the university administration are compared with self-reported grades in a student survey.

Sources of auxiliary data are rich sampling frames, external records, paradata, and for face-to-face-interviews detailed interviewer observation. Sometimes the sampling frame contains rich information on individual level for both respondents and nonrespondents. For instance, when a sample of students is drawn from a university data base and administrative information on students is available. National Statistical Institutes that are allowed to use the population registers, which contain demographic background variables, also have access to rich sampling frames.

Need for Auxiliary Data to Estimate Mode Effects

When investigating mode effects, the situation becomes more complicated. There are two types of mode

effects: mode selection effects and mode measurement effects. Mode selection effects are *desirable* effects; researchers use a second mode to get in those respondents that they have missed with the first mode. If a second or third mode just brings in more of the same respondents, researchers could as well stick to a single-mode study and preferably use the least expensive data collection mode. However, mode measurement effects are *undesirable* effects. Mode measurement effects occur, not because a respondent in reality thinks differently, but just because the question is asked in a certain mode. Prime examples are increased social desirability in interviews than in self-administered questionnaires and more context effects in self-administered matrix questions than in sequentially posed interview questions. To separate wanted mode selection effects from unwanted mode measurement effects and in a later stage correct for the unwanted mode measurement effects, auxiliary data are needed.

Demographic data, from sampling frames or registers, can be used to separate mode selection and mode measurement effects. This is under the assumption that the demographic variables used are mode insensitive, that is, unaffected by mode measurement effects.

A second source of auxiliary data can be a single mode reference survey. This reference survey is seen as the gold standard, and it is assumed that the reference sample is representative of the same population as the mixed-mode study and that there is measurement equivalence. This means that measurement error in the single-mode reference survey and the comparable mode in the mixed-mode study is equal.

Existing data sets are generally used when utilizing biographical data or a reference survey. In some cases, it is possible to collect new data as auxiliary data needed for the analysis stage.

When investigating measurement error, it is possible to embed a randomized experiment in the data collection. A subsample is then randomly assigned to the two modes, and the resulting data are used to investigate and calibrate for measurement error. A second method is the reinterview, again a subsample is reinterviewed in a single mode. This reinterview then acts as a benchmark.

To estimate nonresponse, selection, and measurement error, researchers need auxiliary data. Whatever the source of these additional data is, researchers have to pay attention to the following points:

1. Which data serve as benchmark (gold standard) and why?
2. What is the population investigated to collect the auxiliary data and which population is investigated in the (mixed-mode) survey?
3. Do these population differ and on what points?

Only then can a methodologically sound analysis be performed.

Analysis

This section discusses weighting in single- and mixed-mode surveys. Different adjustment methods are

discussed, as well as their pros and cons.

Weighting in Single-Mode Surveys

As mentioned earlier, to correct for selection effects, researchers can compare demographics for the sample obtained with known demographics for the population and use weighting to make the demographics of the sample equal to those of the population. There are a variety of methods to calculate the weights (see Biemer & Christ, 2008). The simplest weights, base weights, are the inverse of the response probabilities of the sample units. For most surveys, these weights are easy to calculate. More complicated weighting schemes combine design weights, selection weights, and poststratification weights. If response information and covariates (e.g., demographics) are available for both the respondents and nonrespondents, logistic regression can be applied to estimate the propensity to respond. The weights are then based on the inverse of the response propensities (propensity score weighting).

Weighting comes with a set of assumptions and analysis issues. One is that weighting assumes that covariates are used that strongly predict the response probabilities. Demographics are usually only weakly related to the observed response probabilities, and there will probably be unobserved variables that are better predictors. If these unobserved variables also correlate with the target variables, then the researcher has biased estimates, and weighting on demographics will not remove these biases. Another issue is that weighting increases the sampling variance, so in the analysis of weighted data, specialized software is needed to accommodate this.

Weighting in Mixed-Mode Surveys

In mixed-mode surveys, one way to accommodate mode differences is to weigh the subsamples in the different mode individually to reflect the population demographics. This method is easily recognized as naive. One important reason to carry out a mixed-mode survey is that different modes may appeal to different subsections of the population, so actually different groups of respondents are wanted. Therefore, weighting these subsections toward an average population distribution actually makes no sense. More sophisticated correction methods that distinguish between the wanted selection and the unwanted mode measurement effects are needed. A selection of such methods is introduced in the following subsections.

Distinguishing Between Selection and Measurement Effects

As discussed earlier, mixed-mode surveys may have both positive and negative effects on data quality. For example, to the extent that different modes attract different respondents, mixed-mode surveys will lead to better coverage, which is one of the reasons to carry out a mixed-mode survey. This mode selection effect is generally considered a beneficial effect. However, there may also be mode measurement effects on the response process, which leads to differential errors of observation. An example discussed earlier is social desirability effects, which are known to be greater in survey modes that involve an interviewer. A different mode measurement effect is changes in the question–response process that may lead to measurements that

are not equivalent (or invariant) across different modes.

The observed mode effect is the combination of the mode selection and the mode measurement effects, which are confounded. Thus, the problem in mixed-mode surveys is to distinguish between selection and measurement effects, and adjustments should be done only for measurement effects, not on selection effects.

Like weighting adjustment, adjustments for mode measurement effects rely strongly on auxiliary data. The following subsections describe four approaches to adjusting for mode measurement effects: covariate adjustment, counterfactuals or potential outcomes, using a reference survey, and reinterviewing.

Covariate Adjustment

Covariate adjustment basically uses analysis of covariance to correct for covariates that explain differences between the modes used in a survey. There are two varieties: front door and back door (Vannieuwenhuyze, 2014). The front door approach uses covariates that predict the selection effect and interprets differences found after covariate adjustment as mode measurement effects. It derives its name from the fact that it uses covariates, usually demographics, to predict participation in different modes, before the target variables are analyzed. The back door approach uses covariates that predict the measurement effect, its name derives from using the target variables after the data are collected. Both approaches come with strong assumptions. The front door approach assumes that the covariates fully explain the selection effect and that the covariates are not themselves affected by a mode measurement effect. With demographics such as gender and age, the second assumption is justifiable, but the first assumption is unlikely to be met. The back door approach assumes that the covariates explain the mode measurement fully and that there is no selection between modes on the covariates. Here, both assumptions are unlikely to be met. Note that the front door approach uses the same information as the analysis of covariance approach. Analysis of covariance assumes linear relationships, and weighting does not. Thus, by requiring less strong assumptions, weighting has a methodological advantage.

Potential Outcomes

The counterfactual or potential outcomes approach poses the counterfactual situation that all respondents can be surveyed in all available modes. Thus, in each mode, there are potential outcomes, while in practice each respondent is surveyed in one single mode only. The potential outcomes in the other modes are all missing. Modern missing data techniques are then used to analyze the incomplete data.

Like the other approaches, the potential outcomes approach requires good auxiliary data. In incomplete data analysis, two kinds of auxiliary data are important: variables that predict the missingness, and variables that predict the unobserved values. Typically, multiple imputation is used to impute the potential outcomes in the unobserved modes, and the data are analyzed as if all respondents were indeed measured in all modes.

In the potential outcomes approach, there is no distinction between diagnosis and adjustment; after the imputation procedure for all subjects all target variables are observed in all modes. Differences between the

modes can directly be analyzed to investigate mode effects. In addition, substantive analyses can be carried out, either combining all modes used or choosing one of the modes as a preferred mode and carrying out the substantive analysis only in that mode.

Multiple imputation can be implemented in several ways. Parametric imputation techniques rely on a model to produce the imputed values. The relevant assumptions are that the model fits the data well and that the distributional assumptions of the model are met. If the sample is large, as it generally is in surveys, nonparametric or semiparametric methods are preferred. These techniques match cases with missing values to cases that have observed data for these variables (donors), and which are similar to the recipient. These techniques differ in the way they define similarity. In large samples, there are usually many possible donors for each recipient, and a random procedure picks different donors in different imputed data sets. The advantage of nonparametric imputation is that, if the target variable has a non-normal distribution, the imputed values tend to reflect that distribution. In addition, the procedure always imputes missing data with values that have been observed in the data. Given that the potential outcomes approach leads to data with a large fraction of missing values, the advice is to generate a large number of imputed data sets (50 to 100) and to monitor the imputation process carefully.

The Reference Survey

A very different approach is to use data from a reference survey as benchmark. The reference survey is a survey of the same population, with the same survey questions, using only a single mode. The reference survey is often a face-to-face survey because that tends to have the highest response rate, but this is not a requirement. The reference survey can be embedded in the survey design by assigning a random subsample of the respondents to a single mode. Jorre Vannieuwenhuyze and Geert Loosveldt (2013) discuss several ways to analyze the combined mixed-mode plus reference survey data. The most general model uses latent class analysis. The mixed-mode sample contains respondents who have been observed in different modes. In this sample, the variable “mode of observation” is an observed categorical variable. In the single-mode sample, the variable “mode of observation” is unobserved or latent because the respondents were not given that choice. The mixed-mode and reference sample are combined, and a latent class variable is specified that is observed in part of the combined sample. The latent class variable can be predicted by background variables, but it can also be defined by the model specified for the target variables. That model can be as simple as the estimation of a mean for a target variable, but it can also be a regression or structural equation model (SEM). The advantage of this approach is that comparisons between modes can be explicitly made as a multigroup model in SEM. The question of measurement equivalence or model equivalence across different modes can be tested empirically using well-established SEM approaches.

Reinterviewing Respondents

A form of reference survey is reinterviewing a subsample of the respondents in a second single-mode survey, using the same or a shortened version of the survey questionnaire. This assumes that memory effects are

negligible and that there is no true change between the two measurement occasions. This means that the time period between the two measurement occasions must be carefully chosen, taking into account the actual survey questions and the intended population. Since the repeated data in the single mode reinterview is from the same respondents, these single mode-data can be used to adjust the original mixed-mode survey data.

Conclusions for Adjustment in Mixed-Mode Surveys

One important conclusion, already presented in the design section, is that auxiliary data are of paramount importance. Collecting such data should be part of the design and implementation of a mixed-mode survey, and the design of the survey questionnaire and procedures should aim at minimizing mode measurement effects. But auxiliary information is needed to distinguish between the desirable selection effects and the undesirable mode measurement effects.

As the discussion of the various adjustment techniques shows, the analysis can become quite complex. With repeated surveys, researchers may end up with a survey design that is so finely tuned that diagnosis and adjustment of mode differences is no longer needed. But the burden of proof for robust measurement equivalence across survey modes is on the shoulders of the researchers.

Finally, Joop Hox, Edith de Leeuw, and Thomas Klausch (2017) discuss design and analysis issues in mixed-mode research. As an example, they use a data set from a panel, where a mixed-mode data collection was preceded with a single-mode data collection from a previous wave of the same survey. This previous data collection wave was considered as rich auxiliary data, which makes strong adjustments possible. With these data as example, they discuss analysis of covariance, multigroup SEM, potential outcomes and the reference survey approach. Reassuringly, this empirical example indicates small mode measurement effects, and more reassuringly, although they find some differences between the various approaches, all approaches lead to fairly similar conclusions. Given that some of the analysis techniques presented are complex, survey practitioners might wish to choose an adjustment approach with which they are familiar or for which they can obtain good advice.

Future of Survey Research: From Single to Integrated Research

An important problem with single-mode research and TSE are response rates, but mixed-mode research is difficult due to increased complexity in logistics and the difficulty of separating selection from measurement effects. In addition, with the rise of data science and big data, many more opportunities for data collection are available to researchers. Here, the trends in single- and mixed-mode research are discussed first, before discussing emerging technologies leading the way to integrated research.

The online survey is the most popular mode for single-mode surveys in most countries. Many online surveys

use online panels since repeatedly asking the same respondents is cost-effective. In addition, many use probability-based panels (often recruited off-line) to warrant data quality. With the rise of the mobile device, surveys have been redesigned to make them mobile-friendly, taking into account small screen sizes and touchscreen as method of navigation. Mixed-device surveys, as online surveys are being completed on a range of different devices, have many similarities to mixed-mode surveys since the differences between devices can affect TSE in a similar way as mixed-mode surveys. The face-to-face survey has become too expensive for most types of research. In addition, telephone surveys are rarely being used in most countries due to low response rates (often below 5%), making telephone surveys very expensive as well. As a follow-up mode in an existing relationship (hot contact), telephone surveys can still be effective in increasing response rates to mixed-mode surveys.

With low response rates for self-administered surveys and high costs for interviewer-administered surveys, there is a tendency for more mixed-mode surveys. For example, National Statistical Offices rely mostly on mixed-mode surveys to produce official statistics, and mixed-mode projects are funded by EUROSTAT. Countries differ, however, in the combination of types of mode. For example, the United States often combines online and mail surveys, while (Western) European countries often combine online and interviewer-administered surveys.

Another trend in data collection is data fusion, whereby different sources of data are integrated with survey data. Big data are often freely available, but surveys are still needed to answer the “why do people do this” question. “What” people do can then be estimated based on big data such as social media, sensor, transaction, and register data.

Apart from using big data, which is not collected specifically for research purposes, there are other types of big data that can be collected for research purposes such as passive measurements. The 21st century has seen a vast increase in the use of different types of devices. Tablets, smartphones, but also smartwatches, and other wearables such as fitbits, make it possible to measure a range of activities. All these devices can be linked together, creating a vast array of possibilities by way of passive measurement. Sensors in mobile phones can measure location and movement. In addition, the camera can take photos or videos. All this information can be used in addition to survey data or to replace survey data. Cumbersome surveys such as surveys about travel, time budget studies, or budget surveys, do not have to rely on lengthy survey questions with measurement error due to recall error or omissions. The camera can be used to scan receipts or take a picture and these photos can be uploaded without any survey question being asked. In addition, tracking respondents via GPS on the mobile phone can replace travel diaries. TSE should be taken into account for all forms of passive measurement. Many people are not willing to give consent to track their GPS coordinates or are not willing to install an app to scan receipts or take pictures. Although one typically expects less measurement biases in passive measurement (although this is an assumption), selection biases often have a significant effect on total (survey) error.

With the popularization of the Internet, many researchers have undertaken online surveys, but other modes such as face-to-face surveys, telephone surveys, and paper mail surveys are still being conducted today.

With the use of big data and the creation of applications, sensors in phones and other devices, some survey questions can be replaced by using big data or passive measurement in devices, but there will still be a need to use traditional forms of survey research since many research questions cannot be answered by big data or passive measurement. Optimizing survey questions remains as important as developing tools for passive measurement. The possibilities for conducting survey research grow in a rapid pace, and careful consideration of mixing modes, applications, and data sources is essential in estimating total (survey) error and the accuracy of the data being used. Integrated research, by combining different forms of data, will be the future though and is broader than merely combining survey modes.

Further Reading

de Leeuw, E. D. (2018). Mixed-mode: Past, present, and future. *Survey Research Methods*, 12, 75–89. Retrieved from https://ojs.ub.uni-konstanz.de/srm/article/view/7402/6582

de Leeuw, E. D., & **Berzelak, N.** (2016). Survey mode or survey modes? In **C.Wolf**, **D.Joye**, **T. W.Smith**, & **Y.-C.Fu** (Eds.), *The SAGE handbook of survey methodology* (pp. 142–156). Thousand Oaks, CA: SAGE.

de Leeuw, E. D., **Hox, J. J.**, & **Dillman, D. A.** (2008). *International handbook of survey methodology*. New York, NY: Lawrence Erlbaum/Psychology Press, Taylor and Francis Group.

de Leeuw, E. D., & **Toepoel, V.** (2018). Mixed-mode and mixed-device surveys. In **D. L.Vannette** & **J. A.Krosnick** (Eds.), *The Palgrave handbook of survey research* (Chapter 10, pp. 51–61). Cham, Switzerland: Palgrave.

HoxJ., **de Leeuw, E. D.**, & **Klausch, T.** (2017). Mixed mode research: Issues in design and analysis. In **P. P.Biemer**, **E. D.de Leeuw**, **S.Eckman**, **B.Edwards**, **F.Kreuter**, **L. E.Lyberg**, **C.Tucker**, & **B.West** (Eds.), *Wiley Series in Survey Methodology: Total survey error in practice* (Chapter 24). New York, NY: Wiley.

Summaries, glossary and additional material. Retrieved from https://edithl.home.xs4all.nl/surveyhandbook/contents.htm

References

Biemer, P. P., & **Christ, S. L.** (2008). Weighting survey data. In **E. D.de Leeuw**, **J. J.Hox**, & **D. A.Dillman** (Eds.), *International handbook of survey methodology* (pp. 317–341). New York, NY: Taylor & Francis.

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21, 233–255. Open access, Retrieved from https://dspace.library.uu.nl/handle/1874/202577

de Leeuw, E. D. (2018). Mixed-mode: Past, present, and future. *Survey Research Methods*, 12, 75–89. Retrieved from https://ojs.ub.uni-konstanz.de/srm/article/view/7402/6582

de Leeuw, E. D., & **Hox, J. J.** (2011). Internet surveys as part of a mixed-mode design. In **M.Das**, **P.Ester**, & **L.Kaczmirek** (Eds.), *Social behavioral research and the internet* (pp. 45–76). New York, NY: Taylor & Francis.

de Leeuw, E. D., **Hox, J. J.**, & **Dillman, D. A.** (2008). *International handbook of survey methodology*. New York, NY: Psychology Press.

Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology*, 43. Retrieved from http://www.statcan.gc.ca/pub/12-001-x/2017001/article/14836-eng.htm

Groves, R. M., **Fowler, F. J., Jr.**, **Couper, M. P.**, **Lepkowski, J. M.**, **Singer, E.**, & **Tourangeau, R.** (2009). *Survey methodology*. Hoboken, NJ: Wiley.

Groves, R., & **Lyberg, L.** (2010). Total survey error. Past, present and future. *Public Opinion Quarterly*, 74, 849–879.

Hox, J. J., **de Leeuw, E. D.**, & **Klausch, T.** (2017). Mixed mode research: Issues in design and analysis. In **P. P.Biemer**, **E. D.de Leeuw**, **S.Eckman**, **B.Edwards**, **F.Kreuter**, **L. E.Lyberg**, **N. C.Tucker**, & **B.West**. (Eds.), *Total survey error in practice* (pp. 511–530). Hoboken, NJ: John Wiley.

Lohr, S. (2008). Coverage and sampling. In **E. D.de Leeuw**, **J. J.Hox**, & **D. A.Dillman** (Eds.), *International handbook of survey methodology* (Chapter 6, pp. 97–112). New York, NY: Psychology Press.

Lyberg, L. E., & **Weisberg, H. F.** (2018). Total survey error: A paradigm for survey methodology. In **C.Wolf**, **D.Joye**, **T. W.Smith**, & **Y. C.Fu** (Eds.), *The SAGE handbook of survey methodology* (pp. 27–40). Thousand Oaks, CA: SAGE.

Medway, R. L., & **Fulton, J.** (2012). When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly*, 76, 733–746.

Scheuren, F. (2004). *What is a survey*. Retrieved from https://www.whatisasurvey.info/

Tourangeau, R. (2017). Mixing modes: Trade-offs among coverage, nonresponse, and measurement error. In **P.Biemer**, **E.de Leeuw**, **S.Eckman**, **B.Edwards**, & **F.Kreuter** (Eds.), *Total survey error in practice* (Chapter 6). Hoboken, NJ: Wiley.

Vannieuwenhuyze, J. T. A. (2014). On the relative advantage of mixed-mode versus single-mode surveys. *Survey Research Methods*, 8, 31–42.

Vannieuwenhuyze, J. T. A., & **Loosveldt, G.** (2013). Evaluating relative mode effects in mixed-mode surveys—Three methods to disentangle selection and measurement effects. *Sociological Methods & Research*, 42, 82–104.