

Julie de Jong, 2016

(2010 Version: Beth-Ellen Pennell, Rachel Levenstein, and Hyun Jung Lee)

### Introduction Guidelines

1. Before beginning fieldwork, assess the feasibility of conducting the research in each target country and
2. Decide whether the desired information can best be collected by combining qualitative methods with the
- standardized survey.
3. Reduce the potential for nonresponse bias as much as possible.
4. Time data collection activities appropriately.
5. Institute and follow appropriate quality control measures.
6. Document data collection activities.
7. When possible, conduct validation studies to estimate bias.

### References

**Appendices:** [A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#)

### Further Reading

**Introduction** Collecting comparable data in the context of multinational, multiregional, and multicultural ('3MC') surveys is a highly complex task in which one can expect to encounter a variety of languages and cultural contexts. Even within a single country, the target population may not be linguistically, ethnically, or culturally homogeneous. Such cultural heterogeneity could manifest itself through a wide variety of dimensions that could impact data collection efforts. For example, languages spoken may not have a standard written form, or respondent literacy rates may be vastly different. Geographic topography may be difficult (e.g., remote islands, deserts, or mountainous regions), and weather and sea impediments (e.g., winter/summer, monsoons), national and religious holidays (e.g., the Christmas season, Ramadan), and political upheavals may make the harmonization of fielding times across different countries impractical. Moreover, populations may be inaccessible because of migration patterns or interviewer safety concerns, or they may only be accessible under special circumstances (e.g., miners in camps, or populations in which part of the population goes on hunting or fishing trips). Countries also vary widely in both their survey research infrastructures and in their laws, norms, values, and customs pertaining to data collection and data access. Certain modes of administration may not be appropriate or feasible in some situations. In addition, the size and composition of nonresponse will likely vary due to differences in cooperation and ability to contact respondents. Some countries officially prohibit survey research (e.g., North Korea) and severely restrict data collection on some topics.

While a survey conducted in a single country might face one or more of the challenges mentioned above, the probability of encountering multiple hurdles is much higher in a large-scale 3MC study. What is atypical in the one-country context becomes the norm in 3MC contexts. Moreover, the assumed homogeneity and common ground that may, broadly speaking, hold for a single-country study contrasts with the obvious heterogeneity of populations, languages, and contexts encountered in multinational studies. Because of the heterogeneity of target populations in cross-cultural surveys, all some flexibility in data collection protocols can reduce costs and error.

In some cases, a coordinating center dictates data collection decisions across all countries involved. The European Social Survey (ESS), for example, mandates the mode in each country, while the International Social Survey Programme (ISSP) allows a certain amount of flexibility. See [Study Design and Organizational Structure](#) for more details. These guidelines are intended to advise data collection decision-makers as they consider the issues and requirements relevant to different modes of data collection and provide extensive recommendations for the practical implementation of said modes. Because guidelines and lessons learned vary greatly depending on the specific mode of data collection, we begin with general considerations relevant for data collection in any mode, and then provide further guidelines and lessons learned in the subsequent subchapters for the main modes of data collection used for 3MC surveys as follows: Data Collection: General Considerations (these guidelines) [Data Collection: Face-to-Face Surveys](#) [Data Collection: Telephone Surveys](#) [Data Collection: Self-Administered Surveys](#) For a discussion of the advantages and disadvantages of specific modes, the factors involved in mode choice, and whether to standardize mode across locations, see [Study Design and Organizational Structure](#)

[Structure](#). Because difficulties in data collection can be extreme in countries where infrastructure is limited, these guidelines heavily emphasize the challenges of data collection in such contexts.

[↑ Back](#)

**Guidelines Goal:** To achieve an optimal cross-cultural data collection design by maximizing the amount of information obtained per monetary unit spent within the allotted time, while meeting the specified level of precision and producing comparable results.

[↑ Back](#)

## **1. Before beginning fieldwork, assess the feasibility of conducting the research in each target country and culture**

### ***Rationale***

Local knowledge can be critical to understanding cultural traditions and customs, possible limitations, and the feasibility of the research. Experienced researchers, interviewers, and key stakeholders familiar with the topic or population under study can help assess concerns and challenges and suggest potential solutions.

### ***Procedural steps***

- 1.1 Assess the appropriateness of (1) the constructs to be studied and (2) the mode of data collection selected. For detailed information about different data collection modes, see [Data Collection: Face-to-Face Surveys](#), [Data Collection: Telephone Surveys](#), and [Data Collection: Self-Administered Surveys](#).
- 1.2 Gather information from the coordinating center on major survey design features. These might include the survey topic and questionnaire items, intended mode of administration, instrument technical design, respondent burden, length of interview, complexity of topic, and proposed methods for dealing with nonresponse.
- 1.3 Gather information from people in the area who are familiar with data collection as well as who may not be familiar with survey data collection, are familiar with, represent, or may share characteristics with the population of interest. If possible, conduct focus groups and one-on-one interviews with individuals within the contracted survey organization and others who have previously collected data within the country or location.
  - 1.3.1 Solicit the help of local collaborators or researchers. Local collaborators may have a solid understanding of relevant cultural concerns or challenges, or they may be able to help gather information from other local individuals who are more familiar with data collection and the population(s) of interest.
    - Provide local collaborators or researchers with a detailed description of the protocol, including the proposed mode of data collection; nonresponse reduction techniques; timing; interviewer training, remuneration, monitoring; and the general framework for data collection.
    - Explain and clarify any survey terminologies to ensure common understanding.
    - Request feedback on all aspects of the proposed study.
    - Arrange to be present (even if by phone or other means of communication) when local collaborators are collecting information from local resources to clarify and probe when needed. However, before making a decision to join those meetings, assess whether participating in them might make locals uncomfortable about providing information.
  - 1.3.2 Elicit information from these local human resources and any relevant administrative bodies on:
    - Population issues (e.g., local knowledge about the survey, family structure and household listing issues, literacy levels, unwritten languages and cultural norms).

- Logistical issues (e.g., seasonal accessibility, locked dwelling units, secured or dangerous areas, connection issues).
- Issues related to mode choice (see [Study Design and Organizational Structure](#), [Data Collection: Face-to-Face Surveys](#), [Data Collection: Telephone Surveys](#), and [Data Collection: Self-Administered Surveys](#)).
- Issues related to interviewers, if an interviewer-administered mode is used (e.g. availability of interviewers, background, safety concerns).
- Human protection issues (e.g., legal and cultural permissions which may be necessary to conduct the study) (see [Ethical Considerations](#)).

## ***Lessons learned***

- 1.1 While outside input is often helpful, recognize that negative feedback may, in part, reflect uncertainty rather than concrete obstacles. Such feedback can, however, alert researchers to constraints that require attention. For example, in an early survey of mass media communication behavior in the Middle East, experts predicted that data collection would not be possible in Arab countries because they believed that the populace would think that the interviewers were government agents. The experts also suggested that women could not be hired as interviewers and that it would be impossible to survey previously unsurveyed groups, such as the nomadic Bedouin tribes. The research team, however, was successful in their data collection efforts .
- 1.2 A mixed-mode design can introduce error through multiple mechanisms, with the effects being magnified in a survey where comparative data is objective. The following lessons learned speak to research on error and associated adjustments in mixed-mode designs.
  - 1.2.1 While a mixed-mode design can reduce the cost of data collection by allowing for increased flexibility to accommodate local contexts, it may also create an additional layer of complexity and, thus, of the overall cost for the subsequent harmonization of data by coordinating centers. The Gallup World Poll implements a mixed-mode design in which telephone interviewing is used in countries where 80% or more of the target population is covered and face-to-face interviewing is used in countries with lower telephone coverage. The reported costs of telephone surveys are much lower than face-to-face modes , so overall data collection costs are reduced. However, comparability problems due to different modes (phone in one country, face-to-face in another) exist .
  - 1.2.2 In a cross-national context, the impact of mode can be confounded with cultural differences. For example, when the International Social Survey Programme (ISSP) began, the required mode was self-administered. However, low literacy levels in some countries necessitated the use of interviewers. Both response rates and reports from substantive measures differed widely, possibly as a result of differences in mode . Therefore, reported variation between countries on survey estimates may indicate substantive differences or may be due to mode effects and interviewer effects.
  - 1.2.3 We advocate that mode measurement effects need to be considered in three phases of data collection: the collection phase, through the prevention of mode measurement effects; the estimation phase, by consideration of effects arising from both mode selection and mode measurement; and the analysis phase, when adjusting for any bias.
  - 1.2.4 While studies estimating and adjusting for mode effects, particularly in 3MC surveys, are few, findings using ISSP data from Italy, Finland and Norway suggest that the mode differences observed in health status distributions are attributable to mode (self-) selection effects. The results are sustained, i.e., the mode effect is not significant after controlling for available covariates in both logistic regression and the propensity score matching approaches.
- 1.3 When comparing 3MC survey contexts worldwide, we listed the following five dimensions to consider: (1) social and cultural context, (2) political environment, (3) economic climate and infrastructure, (4) physical environment,

research traditions and experience. The authors point out that new technologies have proven themselves to not improve and facilitate sampling in data-scarce environments and data collection using alternate modes, but also questionnaire design and interviewer monitoring. In particular, skyrocketing penetration of telecommunication (i.e., cellular phones or improved network coverage), even in remote areas, eases the way for using phone or video surveys in the future.

- 1.3.1 Summarizing various experiences from surveys across sub-Saharan Africa, highlighted sampling and overcoming language barriers as two major strategies to improve data collection processes in the region. Practitioners have been experimenting with innovative technologies including geographic information systems and satellite imagery to efficiently sample rapidly changing urban areas. Additionally, multilingual questionnaire software is an increasingly important tool in the multilingual environments common in the region .
- 1.3.2 Based on their experiences at the Social and Economic Survey Research Institute based at Qatar University, demonstrated how specialized sampling strategies like prespecified sampling based on gender can be more efficient and reduce costs in regions of the world where societal norms require interviewer/respondent gender matching. Evidence from their research also suggests that the nationality of interviewers influences both participation and responses to survey items, with a mismatch between interviewer and respondent leading to lower participation and increased measurement error.
- 1.3.3 highlighted four key issues when conducting survey research in India and China, where surveys tend to be on a massive scale due to the size of the national populations in each country: (1) the complexity of both the organizational structure of the institutions most commonly involved in conducting survey research and the logistics involved in recruiting and managing the requisite number of interviewers; (2) limited availability of geographic information systems for household sampling (especially in China); (3) obtaining permits and approvals from both government and local leaders for survey implementation (in India, especially if there are religious differences between survey teams and local populations); and (4) a multilingual context which leads to on-the-fly translations.

[↑ Back](#)

## **2. Decide whether the desired information can best be collected by combining qualitative methods with the standardized survey.**

### ***Rationale***

A mixed-method data collection approach can increase data quality and validity in a number of ways. Firstly, applying a combination of research methodologies to study the same phenomenon facilitates the validation of findings through cross-verification, while each method counterbalances the potential limitations of the others . Qualitative and quantitative data collection and analysis methods can be used iteratively to strengthen both approaches. For example, unstructured qualitative interviews may permit a more positive interaction between the interviewer and the respondent, increasing the accuracy of the information the respondent provides as well as their willingness to provide such information. Qualitative methods can also place the behavior of respondents into a broader context, and can improve data coding by revealing unanticipated influences. Secondly, mixing qualitative and quantitative methods can address the complexity of sensitive topics or cultural factors more fully than can quantitative methods alone . Finally, it is not necessary to draw a dichotomy between qualitative and quantitative approaches; researchers may remain open to switching between the two called paradigms within the course of a study .

### ***Procedural steps***

Choose data collection methods to fit the aim of the research question .

- 2.1 Consider combining less structured interviewing, field notes, observation, historical materials, or event history calendars with the standardized survey .

2.1.1 In the social sciences, the term 'methodological triangulation' is often used to indicate that more than two methods are used in a study to double- (or triple-)check results (for further information on methodologic triangulation and integrating qualitative and quantitative methods in data collection, see [Further Reading](#))

2.1.2 Triangulation can also widen and deepen one's understanding of the phenomenon being studied.

2.2 Ethnosurveys offer an approach that combines survey and ethnographic data collection and allows each method to inform the other throughout the study. Equal weight or priority is given to each method. Quantitative data is collected in a manner that falls between a highly structured questionnaire and a guided ethnographic conversation, which is helpful in contexts where rigid structure may be inappropriate but where some standardization is needed for comparison purposes. See on the theory and practice of ethnosurveys.

2.2.1 Determine whether your study is retrospective, prospective, or both. Calendar methods are more efficient for retrospective studies, while longitudinal designs are more efficient for prospective studies .

2.2.2 Remember that traditional qualitative methods can be more expensive and time-consuming than a standard survey .

### ***Lessons learned***

3MC projects have successfully combined qualitative and quantitative methods of data collection in many different ways.

2.1 The Tamang Family Research Project, conducted in Nepal in 1987 to 1988, studied two communities to see how family structure influenced fertility decisions. By adding less structured ethnographic interviews to the highly structured survey, the investigators discovered that a previously unknown factor, the Small Farmers Development Program (SFDP), had a significant influence on fertility decisions .

2.2 The event history calendar method is easily adaptable to fit cultural needs. Some tribes in the Chitwan Valley Family Study (CVFS) conducted in Nepal had no conception of time measurement. Researchers successfully used local and national events as landmarks to help respondents accurately recall their life course history .

2.3 To look at trends in household poverty, followed seven steps in a Stages-of-Progress method:

2.3.1 Assembled a "representative community group" (p. 2);

2.3.2 Presented objectives;

2.3.3 Collectively described the construct;

2.3.4 Using current definitions of households as the unit of analysis, inquired about the status of the construct present and 25 years ago;

2.3.5 Assigned households to categories;

2.3.6 Asked about reasons for descent into poverty among a sample of households within each poverty category (relative to previous and current poverty status); and

2.3.7 Interviewed household members.

2.4 believes that health research is best conducted using in-depth interviews, rather than being driven by the questionnaire and preconceived notions. He argues that qualitative methods allow for a more thorough analysis and a holistic understanding of the patients' decision-making processes.



- 2.5 describes the use of mixed methods in the context of country case studies.
- 2.6 Since 1984, the program has provided technical assistance for more than 400 surveys in over 90 countries a world. In 1990, Phase II of the DHS introduced a calendar at the end of one of the core questionnaires to collect information relating to births, pregnancies, terminations, and episodes of contraceptive use. The decision to collect information using a calendar was based on the result of experimental research that showed reporting of informal contraceptive histories in the experimental calendar questionnaire improved the quantity and quality of data collected as well as increasing their analytical potential compared to the standard questionnaire. More information about the DHS contraceptive calendar, its history, and how the data are collected and can be used analytically can be found on the [DHS website](#).
- 2.7 used mixed methods to assess the poverty rankings of individual households in eight villages in rural South Africa. The study aimed to identify the number of poor households and to assess their level of poverty. Working with local researchers, community residents drew a map of their village and located each household within its boundaries. Researchers then asked smaller groups of residents to rank pairs of randomly selected households, asking which household in the pair was poorer and which was better-off. Finally, the responses were coded. The authors found a strong agreement between the subjects' coded perceptions of poverty and a household wealth index generated using statistical methods. Howe and McKay used similar methods to study chronic poverty in Rwanda.
- 2.8 studied the influence of parents and other socialization factors on human development. Working with young adults and their families in Asia, Latin America, Europe, North America, and Africa, she successfully combined qualitative analyses of interviews and participant observation with quantitative analyses of questionnaires and videotape.
- 2.9 Implementing qualitative methods or ethnosurveys helped University of Chicago researcher Douglas Massey gain greater insight into the reasons behind migration in the U.S.
- 2.10 By combining data obtained from both statistical and qualitative analyses, Sampson and Laub were able to more accurately explain and identify changes and consistencies in criminological behavior over a convict's life.
- 2.11 suggest returning briefly to the field when writing the quantitative report for more descriptive information and to explore inconsistencies in the data.
- 2.12 , providing several examples (e.g. from the United States, the Netherlands, Thailand, Sudan, etc.) in which researchers used data produced through the following innovative technologies: remote sensing, Google Maps, crowdsourced data (e.g. Google Earth or OpenStreetMap (OSM)), call data records (CDRs) and GPS data from mobile phones, and Internet and social media data, argue that linking these alternative data can be leveraged to counterbalance the weaknesses of one data source with another. Furthermore, combining alternative technology data sources with traditional surveys may lead to more accurate measures and decrease interview length and associated respondent fatigue.

[↑ Back](#)

### 3. Reduce the potential for nonresponse bias as much as possible.

**Rationale** Optimal data collection maximizes response rates, and thereby decreases the potential for nonresponse bias. 'Nonresponse' refers to when survey measures are not obtained from sampled persons, thereby increasing the nonresponse rate. Nonresponse bias occurs when the people who are nonrespondents differ from respondents systematically. Although the response rate alone does not predict nonresponse bias, a low response rate can be a predictor of the potential for nonresponse bias. Furthermore, response rates have been dropping differentially across countries due to noncontact, increasingly, reluctance to participate. The coordination of a cross-cultural survey can be centralized or decentralized, corresponding focus on either input or output harmonization, as discussed in [Study Design and Organizational Structure](#). These differences in coordination can impact response rates and response bias differentially. For example, in a study using the output harmonization model, where each country uses their own methods and strategies to maximize response rates, nonresponse rates can be calculated and response bias can occur in different ways, whereas in a study using input

harmonization, study countries will be limited in adaptation to local contexts, which in turn also impacts response rate and response bias. See for a more in-depth discussion on nonresponse and nonresponse bias in a cross-national study. For a discussion of nonresponse bias within the survey quality framework, see [Survey Quality](#). **Procedural steps**

### 3.1 Consider the following steps at the community level to reduce nonresponse before beginning data collection

3.1.1 Depending upon cultural norms, gain the support of any 'gatekeepers' (e.g., community leaders or elders) before attempting to reach individual households.

3.1.2 Make all efforts to raise awareness about the need for high-quality surveys, and thus the need for people to take part.

3.1.3 Publicize the survey locally to raise awareness and encourage cooperation.

- If most of the population is literate, consider displaying colorful, attractive leaflets on local bulletin boards in other public spaces.
- Use word-of-mouth channels or local dignitaries (doctors, teachers) as appropriate.

### 3.2 Send pre-notification letters to sampled households, if feasible.

3.2.1 The letter should (1) explain the purpose of the survey, (2) establish the legitimacy of the survey organization and the interviewer, (3) assure confidentiality of answers, (4) notify the household that participation is voluntary, (5) include or announce any gifts or incentives and provide information about them, and (6) provide contact information for the organization (see [Appendix A](#) for an example of pre-notification letters).

3.2.2 There should be a short timespan between the arrival of the letter and first contact by the interviewer; a timespan of several days is ideal. If there is a long delay between the letter and first contact, consider sending a second letter before attempting contact.

3.2.3 Personalize the advance letter with the individual name (if possible and appropriate).

3.2.4 Be aware that survey sponsorship may affect both response rates and the accuracy of the actual data. For example, some respondents may fear repercussions if they do not respond to a survey sponsored by a government agency. While this fear may dramatically increase response rates, the quality of the data may be dubious; respondents may feel that their responses are not genuinely confidential if the survey is sponsored by a government agency, and they may not respond openly. In addition, ethical issues arise in such situations (see [Ethical Considerations](#)).

### 3.3 Nonresponse can be assessed and reduced with effective sample management and interviewer management and monitoring systems and associated paradata. For an in-depth discussion on the use of responsive designs and systems to assess nonresponse and nonresponse bias, see [Paradata and Other Auxiliary Data](#).

3.3.1 Study structure and data collection modes may specify which sample management systems are used. In the case of cultural surveys with strong centralized control, a single sample management system may be specified in the contract with local survey organizations.

3.3.2 A good sample management system facilitates evaluating interviewer workload and performance.

3.3.3 Monitor response rates continuously, and produce reports of daily response rates in order to identify data collection procedures that are more or less successful at increasing participation.

### 3.4 Structure the field staff to aid them in working the sample efficiently and effectively.

- 3.4.1 Give supervisors the responsibility of assigning sample elements to interviewers and reassigning them necessary.
- 3.4.2 Do not allow interviewers to switch sample elements among themselves without the explicit approval supervisor.
- 3.4.3 Ensure that sample elements are assigned in a way that minimizes travel efforts and costs.
- 3.4.4 Decide whether interviewers will work alone, in pairs, or in traveling teams (see above and [Interview Recruitment, Selection, and Training](#)).
- 3.4.5 Decide whether interviewers and respondents should be matched on some characteristic(s) such as gender or ethnicity in order to increase respondent comfort and cooperation. If the respondents' characteristics are not known prior to data collection, develop procedures to make on-the-spot matching possible. For example, to facilitate gender matching, send interviewers into the field in male-female pairs.
- 3.5 Specify the minimum and maximum numbers and the timing of attempts to contact before the final disposition is assigned to increase efficiency.
  - 3.5.1 Interviewers should attempt to contact respondents at different blocks of time across the week to increase the probability of reaching the respondent at home.
    - The times of day when persons are most likely to be at home vary by culture, location, and context. For example, working respondents in the United States are more likely to be reached on evenings and week ends.
    - Alternatively, specify the minimum number of times that attempts must be made during daytime hours, evening hours, and during the weekend (see for details on call scheduling). Incorporate culture-specific information about likely at-home patterns, such as normal workdays, normal work hours, and holidays. Consider the influence of religious and other cultural norms that restrict interviewing at certain times.
- 3.6 If appropriate, offer an incentive for participation .
  - 3.6.1 Adapt the type and amount of the incentive to local customs. Make yourself familiar with country-specific research on incentives.
  - 3.6.2 According to US- and Canada-based research:
    - Present the incentive as a 'token of appreciation' for participating in the survey, not as payment for the respondent's time.
    - Make the token reasonable; it should not be so large that it might raise suspicion about the researcher's organization's motives or be somehow coercive. It should be generally proportionate to the respondent's time and effort.
    - Ideally, provide the incentive prior to the interview. Incentives promised upon the completion of the interview may also increase participation, but to a lesser degree .
  - 3.6.3 Document the use of incentives, including amount and type, time of implementation, and any special considerations such as increasing the amount of the incentive in the final weeks of the study.
    - According to the existing literature, unconditional prepaid incentives seem to be more effective than conditional incentives paid upon completion of the interview . Thus, eliciting feelings of obligation from the unconditional incentive is more effective than rewarding participation.
    - It may be necessary to monitor the extent to which monetary incentives disproportionately encourage the participation of people with low incomes compared to those with high incomes, and thereby have an effect on nonresponse bias. If poorer people are usually underrepresented in the achieved sample, monetary incentives might reduce nonresponse bias, but if poorer people are already overrepresented, incentives might increase nonresponse bias.



- Offering a choice of different types of incentives might attract people from a more diverse background. help to reduce an existing nonresponse bias and counteract the potentially selective effect of offering or specific incentive.
- For financial incentives, interviewers may be asked to record that an incentive was given to a responder similarly, the respondent may need to sign to indicate receipt.
- In deciding whether to use an incentive, weigh the relative time and cost advantages of using an incentive not using one. Incentives may mean less interviewer time in persuading respondents to participate, or less in refusal conversions. The reduction in interviewer time—and thus costs—must be weighed against the providing incentives.
- See [Ethical Considerations](#) for further discussion on the appropriate use of incentives.

3.7 In using a face-to-face or telephone mode, train interviewers to use culturally appropriate reluctance aversion techniques (see [Interviewer Recruitment, Selection, and Training](#)).

3.7.1 Social or psychological factors (e.g., reciprocation, consistency, social validation, authority, scarcity, loss) affects respondents' decision in survey participation. Minimally, train interviewers how to answer anticipated respondent concerns.

3.7.2 Be aware that local customs and legal limitations may prohibit any attempt to recontact someone who declined to participate in the survey. In these cases, using appropriate reluctance aversion techniques become especially important.

3.7.3 Make sure that supervisors monitor interviewers closely on respondent reluctance issues.

3.8 If using a face-to-face or telephone mode, consider assigning supervisors or more experienced interviewers where interviewers have been unsuccessful making contact or achieving cooperation.

3.9 Consider switching modes to increase contact and cooperation.

3.9.1 Some studies in the United States employ a mixed-mode design in which the least expensive mode is used initially, after which time progressively more expensive modes are implemented in order to reduce nonresponse.

3.9.2 Different modes may produce different survey estimates. These mode-specific differences in measurement might be acceptable to the investigator if nonresponse is sufficiently reduced.

3.9.3 If more than one mode is expected to be used and budget permits, examine possible mode effects prior to start of data collection.

- Test for mode effects by administering key questions or questionnaire sections to a randomly split sample of respondents similar to the targeted population (e.g., asking the questions on the telephone for one group and in person for another).
- If it is not possible to test for potential mode effects beforehand, check for differences in responses at the time of data collection.
- Ascertain whether respondents surveyed in each mode produce similar response distributions.

3.10 Have interviewers complete a contact attempt record each time they attempt contact, whether or not the attempt is successful (see [Appendix B](#) for an example of a contact attempt record).

3.10.1 Use disposition codes to describe the outcome of each contact attempt.

3.10.2 Distinguish between (1) completed interviews with eligible persons, (2) non-interviews (eligible persons who refused), (3) non-interviews (unknown if eligible persons), and (4) non-interviews (ineligible persons).

- 3.11 Assign a final disposition code to each sample element in the gross sample at the end of data collection; include new sample elements that may be created or generated during data collection (e.g., for additional family members through half-open intervals).
  - 3.11.1 Provide a clear explanation and training to interviewers before they are allowed to assign final disposition codes.
  - 3.11.2 Take into account that, in some survey organizations, only supervisors can assign final disposition codes.
  - 3.11.3 See [Appendices D, E, F, and G](#) for a description of disposition codes and templates for calculating response rates from the American Association for Public Opinion Research (AAPOR).
  - 3.11.4 See also AAPOR's *Standard Definitions* publication, which also provides definitions for final sample disposition codes and formulas for calculating response, refusal, and other rates, and AAPOR's Response Calculator (available for download [here](#)).
  - 3.11.5 Note that the list of disposition codes may need to be modified for the local situation, and additional codes may need to be defined to account for local conditions.
- 3.12 Minimize the effects of nonresponse bias on analyses as much as possible.
  - 3.12.1 Nonresponse bias is a function of both the response rate and the difference between respondents and nonrespondents on a particular statistic. Because nonresponse bias is statistic-specific, response rates alone do not indicate nonresponse bias. Therefore, estimate the effect of nonresponse bias on key survey estimates where possible (see [Guideline 7](#) below).
  - 3.12.2 If possible, use weighting and imputations (see [Data Processing and Statistical Adjustment](#)).

### ***Lessons learned***

- 3.1 Cross-national differences in response rates can be due to many factors, including differing judgments of interviewers and other local survey staff about the efficacy and subsequent application of particular survey techniques and protocols. A review of response rates from the 1995 round of the International Social Survey Programme (ISSP) found significant differences in response rates, with at least some of the difference likely attributable to mode (face-to-face vs. mail). Even for countries with roughly comparable response rates, source of nonresponse differed, with noncontact contributing substantially to nonresponse in Japan, and refusal contributing substantially to nonresponse in Russia.
- 3.2 Response rates are not necessarily good indicators of nonresponse bias, but nevertheless tend to be used as proxies for bias. In a health study of the elderly in Scotland, healthy individuals were more likely to participate than unhealthy individuals. Because of this difference between the respondents and nonrespondents, the estimate of health was biased even though response rates reached 82% overall.
- 3.3 While the literature has clearly established the positive effects of prepaid and cash incentives upon response rates in minority countries, it is possible that incentives may affect the propensity to respond differently among a population with high rates of poverty. For example, offering a choice of incentives may be more effective at increasing response rates than simply offering a prepaid incentive. Furthermore, in areas with rampant inflation, the value of cash incentives may decrease dramatically within a short period of time.
- 3.4 The same incentive may affect response rates differently across countries or cultures. In the German General Survey (ALLBUS), the same incentive (€10) was offered to all respondents. The authors examined cooperation for Moroccan and Turkish immigrants. The authors found that the incentive affected cooperation differently for

ethnicity and gender: cooperation rates increased as a result of the incentive for Moroccan women, but did not increase for Moroccan men, or Turkish men or women .

- 3.5 The mechanism of incentive efficacy will differ across mode. In telephone surveys, incentives are often sent to the respondent in an advance letter prior to contact, to encourage cooperation; in mail surveys, the incentive may be sent either in advance or along with the mailed questionnaire; and in face-to-face interviews, the respondent generally receives the incentive at the conclusion of the interview; meaning that the actual transfer of the incentive, and therefore its effect on response rate, can differ across mode, leading to further cross-national differentiation in response rates if different countries use different modes in a cross-national survey.
- 3.6 Use caution when choosing to give monetary rewards to study participants. Keller studied the influence of family structure and other socialization factors on human development in Asia, Latin America, Europe, North America, and Africa. Respondents received a cash incentive, and Keller experienced some hostility from families that were not selected for the study (and, thus, not given any monetary rewards) because they did not have young children .
- 3.7 Some studies vary the use of incentives within a country; for example, offering incentives only to respondents in urban areas, where response rates are typically lower; or offering incentives only in cases of refusal, in an attempt to gain cooperation. If considering this approach, be aware of any concerns that might arise from ethics review boards.
- 3.8 Countries have different incentive norms.
  - 3.8.1 For example, in a recent study conducted in Nepal and the United States, respondents in Nepal were highly cooperative and were offered no financial incentive. In the U.S., however, potential respondents were not as cooperative or easy to contact, and incentives were required .
  - 3.8.2 Some 3MC surveys (e.g., the European Social Survey and the ) allow each participating country to decide whether or not to offer incentives.
  - 3.8.3 If incentives are offered, the type may vary from one country to another. For example, the Survey of Health, Ageing and Retirement in Europe (SHARE) offers various incentives depending on the country's culture. Incentives for the World Mental Health Survey vary across participating countries, including but not limited to cash (in the Ukraine and United States), an alarm clock (in Columbia), and a bath towel (in Nigeria); no respondent incentives are offered in Mexico, South Africa, Belgium, Germany, Israel, Japan, or China. In the Netherlands, flowers are a customary gift to the hostess when visiting for dinner; therefore, flowers are an effective incentive in the Netherlands.
- 3.9 Similarly, many cross-cultural surveys (e.g., the European Social Survey, the , and the World Mental Health Survey ) allow participating countries to vary in their use of advance letters and follow-up letters. In the Survey of Health, Ageing and Retirement in Europe (SHARE), advance letters are mailed to each household in the gross sample, and follow-up letters are used with reluctant respondents.
- 3.10 In an experimental design in the U.S., researchers investigated the use of a novel incentive they termed 'reciprocity by proxy,' wherein respondents were invited to participate in a program with the promise that their participation would result in a gift to a third party, such as a charity. Researchers found that reciprocity by proxy increased participation more than either incentive by proxy or no incentive. However, researchers caution that this approach can backfire if the target audience does not support the beneficiary of the gift . To mitigate this risk, researchers can offer to match the contribution to a charity of the respondent's choosing.
- 3.11 An effective sample management system can clarify the causes of nonresponse. When the Amenities and Service Utilization Survey (AVO) was conducted in the Netherlands in 1995, interviewers were not asked to record detailed disposition codes for each call. As a result, refusals could not be distinguished from noncontacts. When the survey was repeated in 1999, detailed disposition codes were collected. Researchers were then able to see that, after three unsuccessful contact attempts, refusal was the more probable explanation .

- 3.12 Not all survey organizations will be familiar with sample management practices. Allow some time in training interviewers to become familiar with the sample management system (see [Interviewer Recruitment](#), [Selection](#) [Training](#)) and check completed forms.
- 3.13 Comparing nonresponse rates and biases in Round 6 (2012/2013) of the ESS, found that response rates differ significantly despite intensive efforts at harmonization of data collection procedures across countries, attributed differences in part to country characteristics which are difficult to quantify, such as the extent to which a country's population may be “over-surveyed.” Such cultural differences require some adaptation of processes across countries even though such differences may conflict with the goals of standardization. Likewise, post-survey adjustments must factor in country-specific circumstances in striving to achieve measurement equivalence.
- 3.14 Evidence from the California Health Interview Survey suggests that cultural factors such as collectivism or proportion of immigrants and foreign language speakers with limited English proficiency can contribute to nonresponse, with clustering occurring at the community level. The authors argue that more refined measures of cultural factors, such as measures of cultural identification or culturally-oriented attitudes, may better explain nonresponse in 3MC contexts.

[↑ Back](#)

#### 4. Time data collection activities appropriately.

**Rationale** A specific survey estimate of interest may determine the timing of data collection activities; for example, about voting behavior will necessarily be timed to occur around an election. Data collection activities may be hampered by inappropriate timing. Face-to-face data collection, for example, may be impossible during a monsoon season, an election or a regional conflict. The guideline assumes that a specific start time and end time to data collection exists; this guideline does not address issues in continuous data collection. **Procedural steps**

- 4.1 Based upon feasibility studies (see [Guideline 1](#) above), evaluate environmental, political, and cultural considerations which might affect the timing of data collection. These could include:
  - 4.1.1 Extreme weather patterns or natural disasters.
  - 4.1.2 War, military or militia rule, or the possibility of hostage-taking.
  - 4.1.3 Religious and secular holidays or migratory patterns of nomadic people. For example, Independence Day (e.g., Bastille Day in France), New Year's Day in China, summer Christmas holiday in Australia and New Zealand, and vacations in July and August in Europe would not be a good time.
- 4.2 Establish a specific start and end date for data collection.
  - 4.2.1 Keep a concurrent fielding period across countries to guarantee cross-national comparability. For example, the ESS requires interviewers across participating countries in Europe to collect data within a four-month period from September to December of the survey year.
  - 4.2.2 If the 3MC project includes countries located in both the northern and southern hemispheres, where summer and winter are in opposition, consider what field period is most feasible for all countries.
  - 4.2.3 Because unexpected events can interfere with data collection activities, remain somewhat flexible to accommodate unexpected events. Include details about any deviations from the anticipated schedule in the study documentation.

#### **Lessons learned**

- 4.1 Coordination of data collection activities across countries or cultures can be difficult, or even impossible. The Afrobarometer measures public opinion in a subset of sub-Saharan African countries. The coordinators for the Afrobarometer note that data collection is especially difficult during national election or referendum campaign seasons, times of famine, and national or religious holidays. Since such events vary across countries and culture, fieldwork activities are spread over a full year .
- 4.2 Timing of data collection activities may be related to the topic of the survey or statistics of interest. The Comparative Study of Election Systems (CSES), for example, studies elections around the world, and therefore times data collection activities according to local election cycles .
- 4.3 The response rate for the Asian Barometer survey in Japan in 2003 was 71%. In 2007, the response rate dropped to 34.3%. One possible reason for the sharp drop in response rates in 2007 is that in 2006, the law no longer allowed commercial surveys to use voter lists or resident registries. As a result, many people mistakenly believed that regulation also applied to academic research .
- 4.4 Data collection in Germany for the first European Social Survey had to be delayed due to general elections that autumn.
- 4.5 In some settings, electrical availability is dictated by the calendar, and should be evaluated prior to data collection. For example, Nepal relies primarily on hydropower, and so electricity shortages increase significantly in most of the country during the dry season between February and April, with some areas being without electricity for more than 14 hours per day. Recharging equipment in these sorts of environments can be a major impediment .

[↑ Back](#)

## 5. Institute and follow appropriate quality control measures.

**Rationale** If errors are caught early, they can be corrected while the study is still in the field, however, improvement during data collection may introduce some measure of inconsistency in the data. This trade-off should be considered if any action is taken . See also [Survey Quality](#) for a discussion of the quality control framework and [Paradata and Other Auxiliary Data](#) for a detailed discussion on using paradata in quality control and survey error reduction. **Procedural**

- 5.1 Evaluate the effectiveness of data collection protocols regularly. Include:
  - 5.1.1 Sample management systems.
  - 5.1.2 Contact protocols.
  - 5.1.3 Reluctance aversion protocols.
- 5.2 With real-time or daily data transmission, quality control routines and error detection can be implemented more efficiently.
  - 5.2.1 The use of technology for data collection allows for collecting and analyzing paradata (such as keystroke timestamps, and GPS coordinates) for monitoring interviewer behavior (if an interviewer-administered mode is used). This allows for early detection of interviewer deviation from interviewing protocol and, therefore, intervention and better data quality. (See '[Lessons learned](#)' below, as well as [Paradata and Other Auxiliary Data](#) for a detailed discussion on using paradata in quality control and survey error reduction. Moreover, post-survey processing time is greatly reduced.
  - 5.2.2 If an interviewer-administered mode is used, observe the interviewers throughout data collection , monitor them more frequently early in the study and less frequently as the study continues.

- 5.3 If an interviewer-administered mode is used, review a random sample of coversheets on an ongoing basis to ensure that correct eligibility and respondent selection procedures are being followed.
- 5.4 If an interviewer-administered mode is used, provide interviewers with feedback, both individually and as a group.
- 5.4.1 Provide immediate individual feedback if there has been a critical error.
- 5.4.2 Provide routine individual feedback for self-improvement.
- 5.4.3 Offer group feedback to focus efforts on improving the process.
- 5.4.4 Continually evaluate the following with respect to interviewers :
- Knowledge of the study objectives.
  - Administration of the survey introduction.
  - Administration of household enumeration and respondent selection procedures.
  - Reluctance aversion efforts.
  - Contact efforts.
  - Rapport with the respondent (e.g., having a professional, confident manner).
  - Standardized interviewing techniques.
  - Data entry procedures.
  - Administrative tasks (e.g., submitting timesheets in a timely fashion).
  - Ability to meet production goals and maintain productivity.
  - Administration of specialized study-specific procedures (e.g., procedures for taking physical measurements, administering tests of physical performance or cognitive ability).
- 5.5 Whenever possible, recontact or reinterview approximately 10–15% of each interviewer's completed cases, at random .
- 5.5.1 If recontacting the respondent, verify that the interview took place, inquire whether the interviewer acted professionally, and ask factual questions (e.g., mode of data collection, interview length, incentive, household composition, and key survey topics) .
- 5.5.2 If reinterviewing the respondent, ask a sample of factual questions that do not have heavily skewed response distributions, were not skipped by many respondents, are scattered throughout the questionnaire, and have answers which are unlikely to have changed between the time of the interview and the verification check .
- 5.5.3 Conduct reinterviews within a time period that is neither so long that respondents will have forgotten the survey nor so short that respondents will remember all the details of the survey .
- 5.5.4 Make sure recontacts and reinterviews are made with the original respondent, and that questions refer to the same time period as was asked about in the original interview .
- 5.5.5 In some countries, it is not possible to perform recontacts or reinterviews due to laws and/or local customs. Document such instances.
- 5.6 If feasible, audio record face-to-face interviews for review.
- 5.6.1 Determine whether cultural norms permit taping.
- 5.6.2 Inform respondents that they may be recorded for quality purposes and allow respondents to refuse to be recorded.



- 5.6.3 Store any tapes safely and securely (see [Ethical Considerations](#)).
- 5.7 Identify potential interviewer falsification.
  - 5.7.1 Implement silent monitoring in centralized facilities, use audio recordings and recontacts in field studies to analyze outliers in the data to detect falsification.
  - 5.7.2 Check responses to stem questions for each interviewer. Questions that have a stem-branch structure—which require specific responses to 'stem' questions—require the interviewer to ask a number of 'branch' questions and are at increased risk for falsification. If a particular interviewer has recorded responses to stem questions that consistently preclude the interviewer from asking the branch questions, the interviewer may be falsifying data.
  - 5.7.3 Examine paradata, such as keystroke data and timestamps, by interviewer to identify potential falsification.
  - 5.7.4 Examine survey data for any duplicate cases, which can indicate falsification as well as data processing errors.
  - 5.7.5 If falsification of data is suspected, contact the respondents involved over the telephone. If respondents cannot be reached via telephone, send out a brief mail questionnaire with a prepaid return envelope.
  - 5.7.6 If falsification of data is suspected, investigate the interviewer's other work and remove the interviewer from data collection activities until the issues have been resolved.
  - 5.7.7 If irregularities or falsified data are discovered, redo the interviewer's cases and delete all of their recorded data.
- 5.8 For approximately 5% of each interviewer's finalized non-interviews, perform random checks with households to verify that ineligibility, refusal, or other status was correctly assigned. Checks may be done by telephone, in person, or by mail, as needed.
- 5.9 If physical measurements are being taken:
  - 5.9.1 Periodically retest the interviewers on the use of any instruments.
  - 5.9.2 Select equipment that can withstand the local conditions (heat, cold, altitude, etc.).
  - 5.9.3 Document the technical specifications of the equipment chosen.
  - 5.9.4 Re-calibrate equipment as needed throughout data collection.
- 5.10 If the survey is being conducted in a centralized telephone facility, follow established monitoring procedures.
  - 5.10.1 Monitor in relatively short (e.g., one-hour) shifts; this is cost-effective and reduces supervisor fatigue.
  - 5.10.2 Use probability sampling to ensure that the number of interviews monitored is proportional to the number of interviewers working each hour (see [Sample Design](#)).
  - 5.10.3 Monitor new interviewers at a higher rate than experienced interviewers.
  - 5.10.4 Select from eligible cases in which the phone is still ringing so that the supervisor is not forced to wait for interviews to begin in order to start monitoring.
- 5.11 Monitor quality indicators consistently throughout the field period; use an electronic system or note them in a log book. Include the following:

- 5.11.1 Distributions of key variables.
- 5.11.2 Hours per interview (HPI), both for the study as a whole and by respondent groups of interest.
- 5.11.3 Number of respondents approached, interviews completed, incomplete interviews, and contact attempts.
- 5.11.4 Response, refusal, and noncontact rates (see [Data Processing and Statistical Adjustment](#)).
- 5.11.5 Outcomes of all contacts and review of disposition code assignment.
- 5.12 Create statistical process control charts (SPCs) to provide timely information on key aspects of the data collection process .
  - 5.12.1 Use the charts to detect observations that are not within predetermined limits (often between one and two standard deviations of the mean).
    - A common use of SPCs in survey organizations is to assess nonresponse reduction methods over the fieldwork period. Using these charts, the impact of interviewer effort on response rates can be easily assessed (see studies in [Survey Quality](#) for additional discussion of SPCs).
  - 5.12.2 Give extreme observations additional attention and try to determine the root cause.
  - 5.12.3 Refer to the charts when deciding whether to release additional sample elements for interviewers to attempt contact, further monitor interviewers, and offer additional training sessions.
- 5.13 Set contact limitations, determining:
  - 5.13.1 The point at which additional attempts to contact a sample element are inefficient.
  - 5.13.2 Whether respondents cooperating after a certain number of contact attempts are significantly different from others on key indicators.

## ***Lessons learned***

- 5.1 Process and progress indicators are often interdependent. Therefore, improving one process or progress indicator may negatively affect another, particularly in the context of attempts to achieve cross-national comparability. For example, the pursuit of higher response rates can actually *increase* nonresponse bias if the techniques used to achieve the higher response rates are more acceptable and effective in some cultures than in others .
- 5.2 In Round 4 of the study, teams of four interviewers traveled together to the field under the leadership of a field supervisor who has at least either an undergraduate degree and experience in collecting data and managing field work teams or no degree but extensive experience. It was the supervisor's job to ensure quality control of survey returns on a daily basis. Interviewers were monitored at all stages, and debriefed daily immediately after interviews. Completed questionnaires were checked for missing data and inconsistencies. Each field supervisor maintained a daily log of observations on sampling and interviewing conditions and political and economic features of the area, and daily telephone reports to headquarters. A fieldwork debriefing was held after all returns were submitted. Sample back-checks were routinely conducted to ensure that the respondent selection was being done correctly. The field supervisor also verified basic information (e.g., respondent age and level of formal education).
- 5.3 The study required all interview teams to travel together under the supervision of a field supervisor and to have a debriefing meeting each evening. Supervisors randomly checked with respondents to make sure the interview was being done properly.

- 5.4 In Round 5 of the , quality control back-checks were performed for at least 10% of respondents and 5% of nonrespondents either in person, by telephone, or by mail. For the respondents, a short interview was conducted to confirm that the interview took place, whether show cards were used, the approximate length of the interview
- 5.5 In the , each field supervisor oversees two interviewers. Each week, the field supervisor observes and evaluates each interviewer's work and documents the process for submission to the national office. Data collection is split into two rounds; the first half of the questionnaire is completed in round one and then checked for accuracy, and the second half of the questionnaire is completed in round two. After the second round, only data entry errors are corrected. Check-up interviews are routinely performed in 15% to 25% of the households.
- 5.6 The Survey of Health, Aging and Retirement in Europe (SHARE) requires all survey agencies to use an electronic sample management system (SMS). All but three participating countries (France, the Netherlands, and Switzerland) use a 'Case Management System' (CMS) developed by CentERdata. This system monitors the survey progress in real time, including screening for eligible respondents, recording contact attempts, ensuring the correct implementation of contact and follow-up strategies, and managing refusal conversion strategies. Bi-weekly reports are generated for the coordinating team.
- 5.7 The recommended supervisor-to-interviewer ratio in the World Mental Health Survey is 1 for every 8 to 10 experienced interviewers, with those countries using a pencil-and-paper mode having a higher ratio than those conducting computer-assisted surveys. Supervision consists of direct observation and/or audio recording of part of the interview for 5% to 10% of each interviewer's work. Supervisors randomly select 10% of interviewed households, confirm the household listings and selection procedure, and repeat some of the questions. Open-ended responses and other quality control checks are reviewed on a daily basis by supervisors, and interviewers recontact respondents to obtain missing data .
- 5.8 Data falsification can be difficult to detect, and there is no one identification strategy. We suggest researchers set a benchmark (in this example, 85%); any two cases where at least 85% of responses are duplicate are flagged as suspicious. However, this strategy has been argued to produce a large number of false positives , and research suggests that each survey has unique parameters that researchers should account for when analyzing data for potential falsification.
- 5.9 In surveys conducted at the Allensbach Institute in Germany, researchers have used two different methods to mitigate interviewer falsification in lieu of recording respondent contact information and performing post-survey verification . In the first method, researchers included a factual question in the survey that asked about a little-known fact that would be unanswerable to most respondents. Later in the survey, a second item provided the information that would answer the earlier factual question. In a valid interview, respondents would not be able to go back in the questionnaire to use this information to answer the first question correctly; therefore, it was expected that the majority of respondents would provide the wrong answer to the first question. However, an interviewer falsifying responses could potentially use the information to correctly answer the first item. Researchers could then identify the interviewer whose respondents had accurate responses for the first survey question and investigate their other completed interviews for a pattern indicating possible falsification. A second technique used by the researchers at Allensbach was to have respondents write responses to open-ended questions. The handwriting could then be examined to see if the interviewer was completing the interviews him or herself.
- 5.10 In order to better understand the extent to which data fabrication occurs, Robbins and Kuriakose developed a program Percent Match, which identifies observations with at least 80% of item-level matching responses with other observations and those deviating from a Gumbel distribution modeling maximum and minimum values in the sample. Testing more than 1000 datasets collected since the 1980s, the authors found that less than 4% of data from OECD countries were flagged for potential data fabrication compared to 26% from non-OECD countries.
- 5.11 In the Latin American Public Opinion Project (LAPOP), geofencing and geotagging of interviews has been implemented for quality control purposes . Using GPS coordinates, an interviewer is alerted if they are outside

requisite PSU, and the anticipated vs. actual interview locations are also audited using a specialized program developed specifically by LAPOP, with interventions in the case of discrepancies.

- 5.12 offer a number of recommendations to increase data quality when using a life history calendar to collect data including use of a standardized approach of interviewing, audio-recording interviews, shortening the feedback period, adjusting domains and interviewer training to cultural differences, and using paradata (e.g., keystroke select particularly problematic interviews or segments.

[↑ Back](#)

## 6. Document data collection activities.

**Rationale** The documentation of data collection procedures is an essential part of the data collection process. Process documentation is necessary for timely intervention. In addition, by understanding what was done in the field, the data are more easily interpreted and understood. **Procedural steps**

- 6.1 Document the following (see [Appendix C](#)):

- 6.1.1 A summary of feedback from the feasibility studies.
- 6.1.2 The interview or data collection process.
- 6.1.3 A description of the mode(s) used.
- 6.1.4 A description of the mode-specific protocols.
- 6.1.5 A description of the sample management system.
- 6.1.6 A description of any paradata collected.
- 6.1.7 Special approaches to reduce nonresponse, including any incentives and nonresponse followup.
- 6.1.8 Outcome rates by key respondent groups, including response, refusal, noncontact, and other nonresponse.
- 6.1.9 Structure of the field staff (e.g., size of interviewer groups and supervisor/interviewer ratio).
- 6.1.10 Timing of the fieldwork for each country or cultural group.
- 6.1.11 A description of quality control procedures and protocols, including:
  - Interviewer monitoring procedures.
  - Outcomes of interviewer monitoring, such as hours per interview and any falsification rates.
- 6.1.12 Any validation study descriptions and outcomes (see [Guideline 7](#) below).

[↑ Back](#)

## 7. When possible, conduct validation studies to estimate bias.

**Rationale** As noted in [Guideline 3](#) above, response rates alone are not good indicators of nonresponse bias; understanding nonresponse bias and making subsequent post-survey adjustments requires information about the nonrespondents. Sampling measurement error bias can only be estimated when 'true' values for survey variables are known or can be modeled (e.g., using latent class analysis). Validation studies can increase confidence in results, assist with post-survey adjustments ([Data Processing and Statistical Adjustment](#)), and address potential criticisms of the study. However, while the inter-

of survey estimates can benefit greatly from validation studies, conducting them may be difficult and prohibitively expensive. Survey methodological experiments are designed up front, and the outcomes are carefully documented. Although these experiments may or may not directly benefit a given study, they are extremely important for the development and building of a body of knowledge in cross-national survey methodology, on which future studies will be able to draw.

### ***Procedural steps***

#### 7.1 Collect data on nonrespondents, if possible, to estimate nonresponse bias.

##### 7.1.1 One approach is to study sample elements that initially refused to be interviewed.

- Draw a random sample of such initial nonrespondents and attempt to interview them under a modified protocol (e.g., increased incentives or a shorter interview).
- This approach assumes that people who were initially reluctant to participate are identical to nonrespondents on key variables; this may or may not be a valid assumption.
- Document the data collection procedures, including the proportion of initial nonrespondents included in the validation study and the mode.

##### 7.1.2 A second approach is to compare respondents and nonrespondents on statistics of interest using information contained in external records (e.g., population register data).

- Complete external records for all sample elements may be difficult to find, inaccurate, or outdated.
- These benchmark data are rarely available for statistics of interest.

##### 7.1.3 A third approach is to calculate response rates within subgroups (e.g., racial, ethnic, or gender groups).

- This approach assumes that subgroup membership is related to the propensity to respond, and assumes that biases in demographic variables are informative of biases in substantive variables.

##### 7.1.4 A fourth approach is to compare estimates to similar estimates generated from outside surveys.

- While estimates similar to estimates from these benchmark surveys can increase credibility, the key survey variables may not exist in the benchmark survey. Furthermore, coverage, nonresponse, and measurement error differences in the benchmark survey are largely unknown.

##### 7.1.5 A fifth approach is to examine the effect of post-survey adjustments on the estimates by comparing unadjusted and adjusted values.

- The use of this approach strongly assumes that the models used to adjust for nonresponse fully capture nonresponse mechanisms at work. While some amount of nonresponse bias may be controlled using the adjustments, they will rarely—if ever—fully control nonresponse bias.
- See [Data Processing and Statistical Adjustment](#) for more information on post-survey adjustments for nonresponse.

#### 7.2 Use methodological studies to assess measurement error.

##### 7.2.1 One approach is to use cognitive laboratory techniques such as cognitive interviews, vignettes, response latency, and behavior coding (see [Pretesting](#)) to assess potential measurement error.

- This approach assumes that laboratory measurements are comparable with those obtained in the survey.
- Many laboratory experiments do not use probability-based samples; therefore, errors detected in the selected laboratory sample may not be representative of errors in the target population.

##### 7.2.2 Another approach is to check outside records for the true value, or a proxy of the true value, of the measure of interest.

- The researcher must have access to the outside records.
- This approach assumes that the outside records are complete and error-free.
- It may be difficult to match the respondent to the outside record.
- Document record collection procedures, including a description of the records and their quality.

7.2.3 A third approach is to embed a randomized experiment within the survey to assess differences in survey estimates among different measurement conditions. In this situation, respondents should be randomly assigned to the experimental conditions (e.g., interview mode).

7.3 Consider using other methods of assessing measurement error.

7.3.1 Reinterview respondents. Reinterviews are especially useful in determining interviewer falsification, also help assess other forms of measurement error (see and for details on estimating simple response variance bias).

7.3.2 Document all aspects of the reinterview procedure, including:

- The respondents who were eligible for the reinterview component of this study (e.g., random 10% of respondents), as well as the total number of respondents selected and how many completed the reinterview.
- The questionnaire used in the reinterview.
- The mode of administration of the reinterview.
- The interviewers who administered the reinterview (e.g., any project interviewing staff, specially designed interviewers, supervisory staff, clinicians, self-administered, etc.).
- The time interval between administration of the main interview and the reinterview (e.g., reinterviews were conducted 1–2 weeks after the main study interview).

7.3.3 Collect paradata that may be correlated with measurement error (e.g., number of keystrokes, length of interview).

7.3.4 Use interpenetration to estimate correlated response variance due to interviewers.

## ***Lessons learned***

7.1 Supplemental studies can be difficult and expensive to implement, but they are useful for validating survey processes. For example, a study of discharged patients at a French hospital found no difference in patient satisfaction rates between early and late respondents. The authors interpreted this finding to indicate that there was little evidence of nonresponse bias in their estimates of patient satisfaction. However, it is unclear if the differences in estimates were due to nonresponse bias or to measurement error.

7.2 Try to use resources to gain knowledge on bias in an efficient way. Validation studies are expensive but can be useful. Therefore, one should first strive for more preventive measures that hopefully make processes almost error-free. Paradata should be collected and analyzed so that processes can improve and display a decreased variability. In some small-scale validation studies, rather than large ones, should be conducted, and used as input to more large-scale improvements of processes and methods. The optimal allocation between the three is unknown, but the general preferred allocation is evident; namely, prevention first, then process adjustments via paradata, and lastly, small-scale validation studies.

[↑ Back to top](#)

## **References**

[↑ Back to top](#)