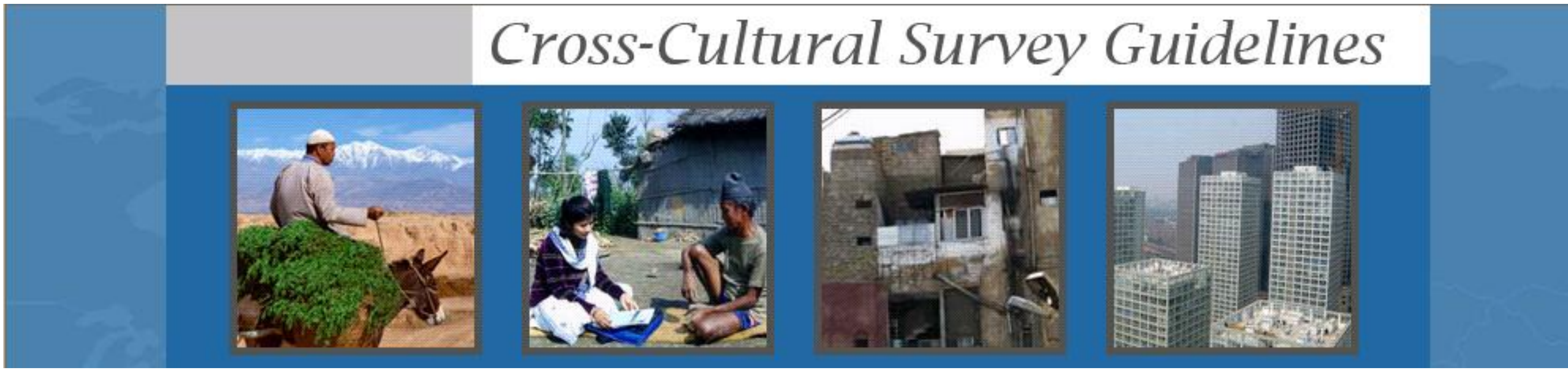
THIRD EDITION

Guidelines for Best Practice in Cross-Cultural Surveys

FULL GUIDELINES



Guidelines for Best Practice in Cross-Cultural Surveys

ISBN 978-0-9828418-1-5 (PDF)

Library of Congress Control Number: 2010937486 2011

Introduction

The number and scope of surveys covering many cultures, languages, nations, or regions have increased significantly over the past decade. This has led to a growing need to provide information on best practices across the multiple phases of cross-cultural survey design and administration to ensure the collection of high quality comparative data. However, there is very little published information on the details of implementing surveys that is specifically designed for comparative research. For example, little has been published on what aspects of cross-cultural surveys need to be standardized and when local adaptation is appropriate. The aim of the Comparative Survey Design and Implementation (CSDI) Guidelines Initiative was to develop and promote internationally recognized guidelines that highlight best practice for the conduct of comparative survey research across cultures and countries. The intended audience is researchers and survey practitioners planning or engaged in cross-cultural or cross-national research. However, we believe that the Guidelines also could benefit researchers and survey practitioners involved in noncomparative survey research.

The goal of the CSDI Initiative has been to develop Cross-Cultural Survey Guidelines (CCSG) as presented here, which cover all aspects of the survey lifecycle. This currently has resulted in 13 chapters. Two additional chapters on survey quality and ethical considerations in surveys are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use, total survey error, and survey production process quality monitoring. This may be affected by survey infrastructure, costs, interviewer and respondent burden, as well as study design specifications. Figure 1 presents a diagram of the survey lifecycle that will reappear in various CCSG chapters, thereby highlighting the chapter's place in the survey lifecycle. The 15 chapters of the CCSG Guidelines are:

I. Study, Organizational, and Operational Structure
II. Survey Quality
III. Ethical Considerations in Surveys
IV. Tenders, Bids, and Contracts
V. Sample Design
VI. Questionnaire Design
VII. Adaptation of Survey Instruments
VIII. Translation
IX. Instrument Technical Design
X. Interviewer Recruitment, Selection, and Training
XI. Pretesting
XII. Data Collection
XIII. Data Harmonization

XIV. Data Processing and Statistical Adjustment
XV. Data Dissemination

Figure 1. The Survey Lifecycle

The CCSG Guidelines draw upon and are based on: (1) general good practice survey methodology, as well as cross-cultural and comparative literature on survey methodology; (2) available study-specific manuals and documentation; and (3) the experiences and lessons learned that authors, reviewers, and editors have added through their work on and with numerous comparative surveys.

Best practices are dynamic and can be expected to evolve over time. At the present time, the Guidelines primarily relate to cross-sectional surveys of households and individuals. At a later point in time, they may be expanded to include establishment and longitudinal surveys. We are also in the process of developing case study examples that demonstrate the application of the Guidelines to the survey lifecycle process, and further elaborating guidelines for assessing the quality of translations.

As more documentation and information about comparative surveys become available, we hope to incorporate the lessons learned from these studies into the CCSG Guidelines. New methodological research will also inform new versions of the CCSG Guidelines. You can greatly help us in these objectives by providing comments and suggestions, or simply alerting us about a topic we need to address. Please contact us at: ccsg_contact@isr.umich.edu.

Citations:

Please cite these Guidelines as follows: Survey Research Center. (2010). *Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved Month, dd, yyyy, from http://www.ccsg.isr.umich.edu/.

The CSDI Guidelines Initiative is led by:

Beth-Ellen Pennell
Kirsten Alcser
Sue Ellen Hansen
Survey Research Center, University of Michigan

Janet Harkness
Survey Research and Methodology Program, University of Nebraska-Lincoln and UNL Gallup Research Center

Photo images:

Some photo images that appear on this Website are stock photos. Others were provided by the Population and Ecology Laboratory in Nepal (Institute for Social Research, University of Michigan), Professor Jintao Xu at the College of Environmental Sciences and Engineering at Peking University, and Yuchieh Lin and the Institute of Social Science Survey at Peking University.

Contributions to these guidelines have been made by:

Authors

Name	Organization	Chapter	Version
Kirsten Alcser	Survey Research Center, University of Michigan (UM-SRC)	Interviewer Recruitment, Selection, and Training; Ethical Considerations in Surveys	2008, 2010, 2011
Christopher Antoun	UM-SRC	Study, Organizational, and Operational Structure; Ethical Considerations in Surveys; Tenders, Bids, and Contracts	2010, 2011
Dorothée Behr	GESIS – Leibniz Institute for the Social Sciences, Survey Design and Methodology	Translators; Translation Budgeting; Translation Scheduling; Translation Tools	2008
Grant Benson	UM-SRC	Survey Quality	2008
Ipek Bilgen	Survey Research and Methodology Program, University of Nebraska-Lincoln, and UNL Gallup Research Center (UNL-SRAM/GRC)	Questionnaire Design; Translators	2008
Emily Blasczyk	UM-SRC	Data Harmonization; Data Processing and Statistical Adjustment; Data Dissemination	2010
Ashley Bowers	UM-SRC	Survey Quality; Ethical Considerations in Surveys	2008, 2010, 2011
Margaret Brill	UM-SRC	Data Collection	2011
Rachel Caspar	Research Triangle Institute (RTI) International	Pretesting	2008, 2010, 2011
AnaLucía Córdova Cazar	UNL-SRAM/GRC	Questionnaire Design; Translators	2010
Judi Clemens	UM-SRC	Ethical Considerations in Surveys; Interviewer Recruitment, Selection, and Training	2008, 2010, 2011
Benjamin Duffey	UM-SRC	Survey Quality	2010
Peter Granda	Inter-University Consortium for Political and Social Research, University of Michigan (UM-ICPSR)	Data Harmonization; Data Dissemination	2008, 2010, 2011

Name	Organization	Chapter	Version
Sue Ellen Hansen	UM-SRC	Survey Quality; Instrument Technical Design	2008, 2010
Janet Harkness	UNL-SRAM/GRC	Questionnaire Design; Adaptation; Translation Assessment; Translation Harmonization; Translators; Translation Budgeting; Translation Scheduling; Translation Tools	2008, 2010, 2011
Lei Huang	UNL-SRAM/GRC	Questionnaire Design; Translators	2010
Frost Hubbard	UM-SRC	Sample Design; Data Processing and Statistical Adjustment	2008, 2010, 2011
Hyun Jung Lee	UM-SRC	Instrument Technical Design; Data Collection	2010
Jennifer Kelley	UM-SRC	Data Collection	2011
Rachel Levenstein	UM-SRC	Data Collection	2008, 2010
Christina Lien	UM-SRC	Ethical Considerations in Surveys	2008
Yuchieh Lin	UM-SRC	Survey Quality; Sample Design	2010
An Liu	UNL-SRAM/GRC	Translators; Translation Budgeting; Translation Tools	2010
Debbie Miller	UNL-SRAM/GRC	Questionnaire Design	2010
Rachel A. Orlowski	UM-SRC	Study, Organizational, and Operational Structure; Tenders, Bids, and Contracts; Data Processing and Statistical Adjustment	2008, 2010, 2011
Beth-Ellen Pennell	UM-SRC	Data Collection; Survey Quality	2008, 2010, 2011
Emilia Peytcheva	RTI International	Pretesting	2008, 2010, 2011
Mathew Stange	UNL-SRAM/GRC	Questionnaire Design; Translators	2010
Ana Villar	UNL-SRAM/GRC; now Stanford University	Questionnaire Design; Translators	2008

Editors

Name	Organization	Version
Lars Lyberg	Statistics Sweden; University of Stockholm	2010

Zeina Mneimneh	UM-SRC	2008, 2010
Peter Mohler	University of Mannheim	2010

Reviewers

Name	Organization	Version
Dorothée Behr	GESIS – Leibniz Institute for the Social Sciences, Survey Design and Methodology	2008
Grant Benson	UM-SRC	2008, 2010
Bill Blyth	TNS-Global	2008
Pam Campanelli	Independent consultant	2008
Somnath Chatterji	World Health Organization, Multi-Country Studies Measurement and Health Information	2008
Rory Fitzgerald	City University, Centre for Comparative Social Surveys	2008
Steven Heeringa	UM-SRC	2008
Achim Koch	GESIS – Leibniz Institute for the Social Sciences, Survey Design and Methodology	2008
Frauke Kreuter	University of Maryland, Joint Program in Survey Methodology	2008
Paul Kussmaul	Retired, formerly Johannes Gutenberg University in Mainz, Germany at Germersheim	2008
Kristin Miller	National Center for Health Statistics	2008
Meinhard Moschner	GESIS – Leibniz Institute for the Social Sciences	2008
José L. Padilla	University of Granada, Social Psychology and Methodology	2008
Alisú Schoua-Glusberg	Research Support Services	2008
Eleanor Singer	UM-SRC	2008
Tom W. Smith	NORC, University of Chicago	2008
Jaré Struwig	Human Sciences Research Council, South Africa	2008
Rick Valliant	University of Maryland, Joint Program in Survey Methodology; UM-SRC	2008
Gordon Willis	National Cancer Institute	2008
Christine Wilson	Heriot-Watt University, School of Management and Languages	2008
Christof Wolf	GESIS – Leibniz Institute for the Social Sciences	2008

Formatting & Copy-Editing

Name	Organization	Version
Gail Arnold	UM-SRC	2008, 2010
Judi Clemens	UM-SRC	2010, 2011
Shaw Hubbard	Independent consultant	2008

Graphic Design

Name	Organization	Version
Larry LaFerte	UM-SRC	2008
Ruth Shamraj	UM-ICPSR	2008

Website Programming

Name	Organization	Version
Tricia Blanchard	UM-SRC	2008, 2010, 2011
Deborah Wilson	UM-SRC	2008, 2010

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Longitudinal study	A study where elements are repeatedly measured over time.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Sample element	A selected unit of the target population that may be eligible or ineligible.

Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.

I I. Study, Organizational, and Operational Structure

Rachel A. Orlowski and Christopher Antoun

Introduction

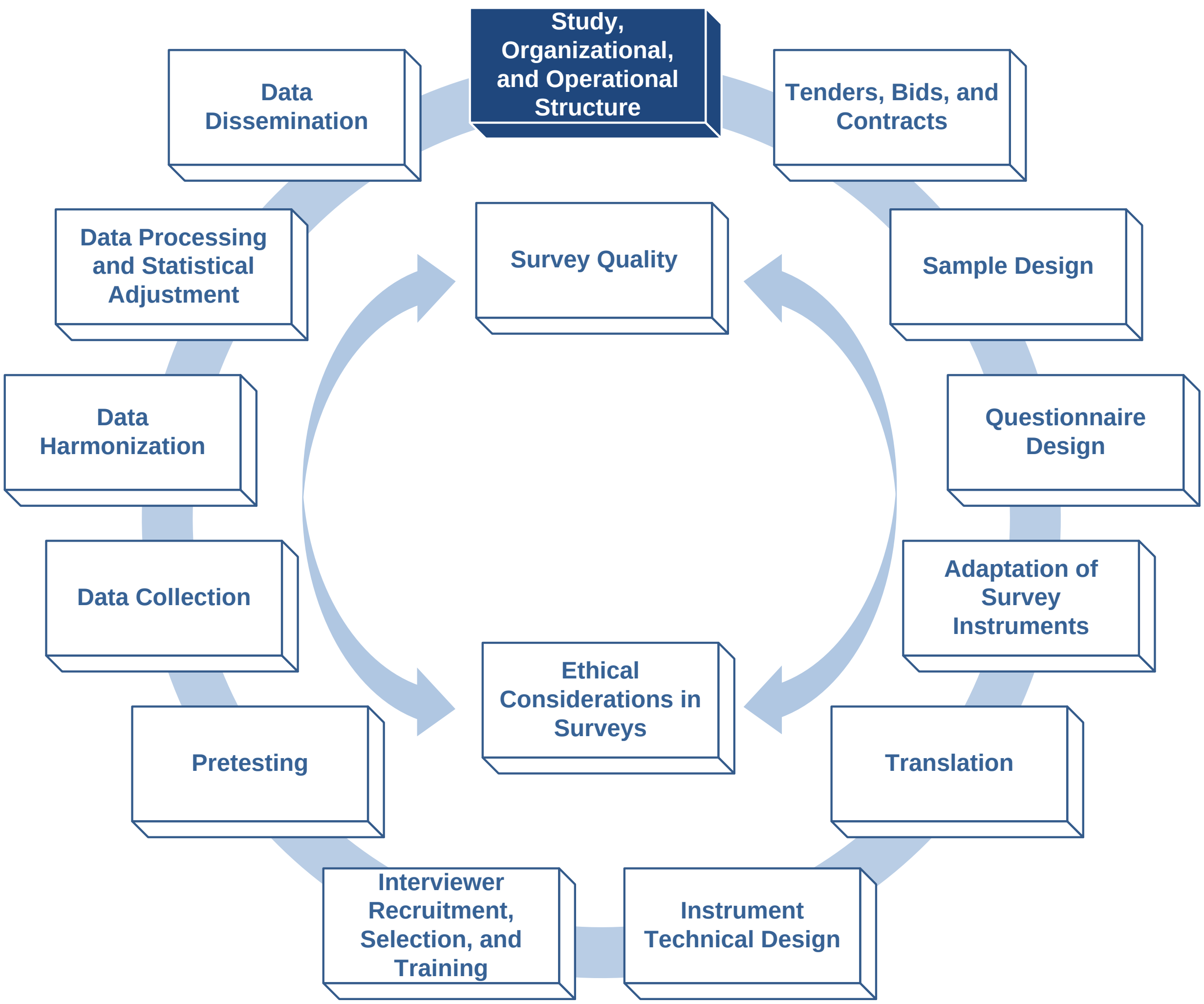
The following guidelines outline a number of study, organizational, and operational considerations which arise when structuring a cross-cultural survey or any survey involving multiple countries, regions, or languages. Several factors will influence how the overall study is designed and later implemented, including the source(s) and flow of funding, the availability of human and technical resources, the best way of contacting and collecting data from respondents, and the research infrastructure. All of these will vary from country to country and culture to culture. Yet, before much time is spent determining the study structure, it is critical to clearly define a study purpose because it drives all subsequent decisions, especially if conflicts between cross-cultural and local interests arise.

Cross-cultural surveys are organized in many different ways, and each has its advantages and disadvantages. These guidelines predominately address a structure with a coordinating center that designs the overall study and assumes the central organizational responsibility to the contracted survey organizations in each country where the study will be carried out. This type of organizing structure is often used in large-scale, cross-cultural surveys. Although not described here, there are situations where the coordinating center is also responsible for data collection in some or all countries. A coordinating center should include people from different countries, institutions, and affiliations. Given this focus, this chapter's primary audience is members of a coordinating center.

With this organizational structure, the coordinating center will specify the operational structure of the survey for each country to follow. It should determine what elements will be standardized across countries and what elements will be localized; there is a balance between standardization of implementation and adaptation to the cultural context. The coordinating center should inform the survey organizations of the quality standards necessary to execute the study.

Figure 1 shows study, organizational, and operational structure within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To establish the study's overall structure and locus of control at all levels and across all aspects of the study's design and implementation, and to communicate this structure to each participating country's survey organization.

1. **Determine the study objectives and identify a study structure that addresses all of the tasks of the survey lifecycle.**

Rationale

Before work is done to organize or operationalize a study, the empirical aims of the research should be understood by all involved. There should be a clear direction and purpose of the research. In order to move the study goals from ideas to a concrete design, a structure of survey tasks should be clearly defined by the coordinating center. This task framework should take into account the cross-cultural nature of the survey.

Procedural steps

- Clearly state the study's objectives, ensuring that central and local study goals do not conflict [2] [6]. When doing so, consider the following main components of design:
  - Representation: What populations are to be studied? (See Sample Design.) [9]
  - Measurement: What data are to be collected? (See Data Collection.) [9]
  - Analysis: What estimates are to be created? (See Data Processing and Statistical Adjustment.) [16]

- Identify tasks necessary to complete all phases of the survey lifecycle.
  - Provide an overview of the possible risks and quality implications associated with every survey task.
  - Consider each subsequent chapter of the Cross-Cultural Survey Guidelines as a potential task in the survey process (also see Appendix A for example considerations for the completion of each task).

- Determine the nature and relationship of tasks. Some tasks tend to have a definite start and end (e.g., sample design), others are ongoing (e.g., ethical considerations), others are often iterative (e.g., questionnaire design), and yet others can overlap (e.g., data collection and data processing). The study structure requires concurrent and coordinated attention to the different tasks in the process [9].

- Evaluate the feasibility of implementing the study given the populations, governments, and politics of the countries being studied, as well as the availability of funding.

Lessons learned

- A failure to communicate overall study goals may lead to local decisions that threaten comparability across countries. For example, a country may remove some locally less salient items from the questionnaire in order to reduce the respondent burden of the interview without realizing that those items are necessary to measure an important survey construct. Conversely, a country may insert items to the questionnaire in order to study a locally-relevant topic without realizing that those items may affect the quality of the data.

- Despite knowing the ideal way of executing a study, the available resources often dictate how a study is structured and implemented. For example, multiple sources of funding are typically needed to provide enough support to coordinate a large-scale, cross-cultural survey; furthermore, each participating country may be funded separately. Funding sources may have requirements that complicate reporting structures within the study and conflict with the goals of the overall cross-cultural survey. The points at issue may relate to a wide variety of features, from data availability to the content of questionnaires. See Appendix B for examples of how existing cross-cultural survey programs have been funded as guidance to ways in which a study can be structured.

2. **Establish an organizational structure for the study at the supranational, national, and, as necessary, subnational levels and define the associated roles and responsibilities.**

Rationale

The coordinating center should first determine its own organizational structure and then set the organizational standards for participating survey organizations. In order to manage a cross-cultural survey efficiently and effectively, roles and responsibilities must be clearly delineated and communicated throughout all levels. This can be accomplished when the central coordinating center works together with local expertise in each participating country.

Procedural steps

- Set up a central coordinating center responsible for managing the overall study and overseeing each country's implementation of the survey.

- Identify the working language for the management of the study.
  - Specify the language proficiency in the chosen working language for all levels of the study management.
  - Do not expect a common understanding of technical terms.

- Become familiar with the culture and political climate of all participating countries in order to establish the most appropriate organizational structure.
  - Review standard handbooks, maps, and ethnographic literature.
  - Review recent media material that depicts the pertinent issues of the participating countries.
  - Identify accommodations that may need to be made at the local level due to (for example) specific legal regulations, local government policies, internet availability, and types of transportation infrastructure (see Data Collection for other considerations).
  - Become knowledgeable about the international, national, and regional business models that affect participating countries [4] [8].
  - Communicate with experienced researchers who have collected data in the participating countries.

- Assess the substantive, methodological, and contractual expertise needed both centrally and for all participating countries. Arrange for expert consultation, as needed.

- Identify the impact that structural aspects of the planned organization have on control, responsibility, and communication at the central and local levels.
  - Determine reporting responsibilities to funding sources.
  - Determine the level of control of the country-specific and cross-national data throughout the study, including analysis and publication of the data (see Data Dissemination).
  - Clearly define subsidiary regulations to specify which decisions are to be made on which level (i.e., supranational, national, or subnational levels).
  - Balance central and local participation in deciding how to address general and country-specific adaptation in processes, methods, and substantive content.

- Consider the creation of a number of task-specific working groups. These groups should be comprised of qualified participants from the participating countries and the coordinating center.
  - Consider creating a working group for each of the tasks mentioned in Appendix A. The responsibilities listed in this appendix could be used as a checklist when making assignments.
  - Specify leadership, authority, and roles across all tasks and levels.

- In addition to working groups, consider the creation of country lead teams to oversee the entire survey implementation in their respective country.
  - Define responsibilities for each lead team member.

- Arrange regular meetings (or conference calls) involving the country lead teams and each working group to discuss study progress.

- Develop a communication flowchart (i.e., who talks to whom about what) with one reference person and one back-up person at each point in the flowchart [20].

Lessons learned

- It is helpful to consider examples of organizational structures when planning a cross-cultural project. See Appendix C for three examples of organizational structures from well-established cross-cultural surveys.

3. Clearly define operational specifications, including deliverables, for each task of the survey lifecycle.

Rationale

Operational specifications ensure that critical aspects of the survey process are defined and then can be controlled. They simultaneously identify required or expected quality standards and help ensure comparability across countries. The specifications should, therefore, be detailed (and measurable, when possible) with clearly delineated deliverables from the participating survey organizations at each task of the survey. In addition, each specification should be justified with a rationale. The specifications form the basis of the country-level tenders and subsequent contracts between the coordinating center and survey organizations (see Tenders, Bids, and Contracts).

Procedural steps

- Taking into account the overall study objectives and weighing the tradeoffs between cost and quality, detail the operational specifications and requirements [2].
  - See the subsequent chapters of the Cross-Cultural Survey Guidelines for procedural steps regarding recommendations for specifications for each task.
  - Before the study is initiated, determine which specifications are more important than others. Communicate to all participants which specifications are a top priority.

- Determine when specifications need to be rigid and when it is possible or preferred to be more flexible.

- Create a study timeline, production milestones, and deliverables with due dates [16].
  - If feasible, use a common project management tool between the coordinating center and each participating country.
  - Keeping in mind local considerations which may affect the study's due dates, require frequent reporting and interim products.
  - Require deliverables with unique identifiers for interviewers and sample elements.

- Consider implementing a system with process checks, using paradata, to recognize when a survey organization is struggling to meet specifications. It is important to identify if a survey organization is having problems meeting specifications as early in the process as possible.
  - Decide what actions to take, as necessary, to rectify delays or delinquencies in meeting specifications.
  - Determine the sanctions/penalties if a participant country continues to fail to meet the specifications.

- Establish a backup plan to ensure the completion of a high quality survey in case countries are unable to meet operational specifications.

Lessons learned

- Adherence to specifications must be controlled. Otherwise, some survey organizations will deviate for a myriad of reasons. Structuring clearly defined specifications and a system of checks and balances will help maintain the highest standards throughout all tasks of a cross-cultural survey.

4. Decide upon quality standards for the implementation of the study.

Rationale

The goal of quality standards is to achieve excellence for all components related to the data [12] [19]. Setting quality standards is critical to ensuring the same level of methodological rigor across countries [6]. Local adaptations will be necessary and appropriate for some aspects of implementation of the study, but any adaptation in the procedure or instrument should be thoroughly discussed, evaluated, and documented beforehand [14]. Frequent measurement and reporting to the coordinating center, along with sufficient methodological support, should allow for timely intervention if problems do arise.

Procedural steps

- Use a Plan-Do-Check-Act cycle (PDCA) by first determining the study's quality standards, then implementing them throughout the research process, while assessing quality indicators at each stage, and finally making appropriate changes to repeat the cycle of PDCA [2] [5].
  - Consider all sources of error in the survey lifecycle, and define quality indicators for key steps in each survey task. See Survey Quality for common sources of error and possible indicators.

- Acquaint study organizers with important quality control literature that distinguishes between common and special causes variation, as well as explains how to act on information about these different kinds of variation [13] [15] [18].

- Form a team in each country that regularly meets to discuss the quality of the local survey. The team should have or should be provided with methodological expertise needed. The team should document and report any concerns to the coordinating center [1] [2].

- Identify tools that control and maintain operational process quality.

- Implement a certification process or a signing-off procedure for each stage in order to check and document that the study design and specification standards are being followed.
  - Quickly address and remedy, if possible, any deviations from expectations that may occur [2].
  - Invoke sanctions, as specified in the contract, if the certification is not fulfilled.

- Consider site visits to all countries to monitor or support the implementation of quality standards. Make sure these visits are specified in the contract with each survey organization.

- Monitor costs in order to avoid overruns.
  - Create a cost-monitoring instrument and checklist.
  - Ensure sufficient funds are allocated to be able to budget quality assessment and documentation activities.
  - Assess risk and determine contingencies for each survey task—weighing cost and errors.

- If and where possible, incorporate methodological research. This will inform long-term quality improvement [11] [19].

Lessons learned

- Variations in country-level research infrastructure, research traditions, and methodological rigor need to be thoroughly investigated and understood when setting quality standards. Some countries will need more assistance in meeting some standards, and this should be taken into account early in the planning process.

5. **Identify and specify the documentation that is required from all levels of participation in the study.**

Rationale

Documentation of procedures ensures that the survey is transparent and allows for replication. Documentation should be detailed and occur throughout the survey lifecycle. If documentation does not occur until the end of the survey or even the end of the survey task, details will likely be lost or forgotten. Therefore, it is important to determine documentation requirements before the study is initiated. The coordinating center should first establish its own documentation procedures and then set documentation procedures for participating survey organizations.

Procedural steps

- Determine documentation procedures for the coordinating center. Require that all decisions regarding the structure be documented, including the study objectives, roles and responsibilities, communication flowchart, and operational specifications.

- Determine the documentation requirements and formats for survey organizations. When appropriate, standardize these requirements across all countries in order to be able to assess quality and comparability. Encourage survey organizations to create clear, concise, and user-friendly descriptions. In addition, these descriptions should be as transparent as possible, with sufficient detail, to ensure that they could, theoretically, be replicated.
  - Determine documentation requirements for all survey implementation procedures. Examples of procedures which should have documentation requirements include:
    - Sampling design and implementation (see Sample Design).
    - Questionnaire development (see Questionnaire Design).
    - Translation of survey materials (see Translation).
    - Mode of data collection decision (see Data Collection).
    - Creation of the production schedule (see Data Collection).
    - Respondent selection and initial contact procedures (see Data Collection).
    - Establishment of supervisory structure (see Data Collection).

    - Data collection observations (recordings, scripted mock interviews, etc.) (see Data Collection).
    - Bio measures collection protocol (see Data Collection).
  - Determine documentation requirements for data collection outcomes (see Tenders, Bids, and Contracts). Detail specifically what is necessary, for example:
    - Interim and final outcome rates (see Data Processing and Statistical Adjustment).
    - Final disposition codes for every released sample element (see Data Processing and Statistical Adjustment).
    - Quality control indicators (see Survey Quality).
  - Decide when survey organizations should share documentation with the coordinating center.

- Record any modifications made either centrally or locally to the study protocol, as well as document the impact of these modifications. Any changes countries make to their protocols and procedures must be carefully documented since these could explain potential differences in the data, either over the course of the study (within a country) or across variables (between countries).

Lessons learned

- Not all deviations that occur in a study can be remedied immediately, but they are helpful for planning future studies. Deviations should be documented to allow for a search of faulty or deficient operational process steps after the fact. This allows for the development of appropriate counteractions at the central and local level for the next wave of surveys.

Appendix A

Survey tasks

When determining the study structure of a cross-cultural survey, it is important that all necessary survey tasks are identified. Below are examples of survey tasks that correspond with each chapter of the Cross-Cultural Survey Guidelines. This appendix provides example considerations for the completion of each task; please see the subsequent chapters for more detailed guidance. By creating a detailed list of survey tasks, the coordinating center can become assured that no aspect of the study structure has been overlooked and can then use this list to assign organizational responsibilities.

- Survey Quality
  - Document the survey process.
  - Develop quality standards and a quality assurance plan.
  - Monitor and support the implementation of quality standards .

- Ethical Considerations in Surveys
  - Create manuals, support documents, and informed consent forms.
  - Observe professional standards and local laws.
  - Ensure the rights of respondents.

- Tenders, Bids, and Contracts
  - Prepare tenders with detailed requirements.
  - Conduct a bidding process and select survey organizations.
  - Negotiate and execute contracts.

- Sample Design
  - Define the target population and determine the sample size.
  - Identify the sampling frame.
  - Implement a selection procedure.

- Questionnaire Design
  - Select a comparative question design approach.
  - Develop protocols for evaluating questions.
  - Adopt questions, adapt questions, and write new questions.

- Adaptation
  - Identify adaptation needs.
  - Modify the questionnaire content, format, or visual presentation.
  - Adapt design features.

- Translation
  - Find, select, and brief translators.
  - Use existing or develop translation tools.
  - Complete language harmonization.

- Instrument Technical Design
  - Develop design specifications for instruments.
  - Develop interface design and programming guidelines .
  - Determine testing specifications.

- Interviewer Recruitment, Selection, and Training
  - Recruit and hire interviewers.
  - Select interviewer trainers.
  - Create a training plan and identify training materials.

- Pretesting
  - Determine the appropriate pretest method and design.
  - Conduct a pilot study.
  - Pretest the survey instrument with the target population.

- Data Collection
  - Select the appropriate mode and develop procedures for that mode.
  - Establish a protocol for managing the survey sample.
  - Manage data collection and quality control.

- Data Harmonization
  - Determine a harmonization strategy.
  - Use a systematic approach to harmonize variables.
  - Compare and integrate information across data files.

- Data Processing and Statistical Adjustment
  - Code survey responses and enter them into electronic form.
  - Edit and clean data.
  - Develop survey weights.

- Data Dissemination
  - Preserve key data and documentation files.
  - Produce public- and restricted-use data files.
  - Prepare final data deliverables and reports.

Appendix B

Funding sources

The source and flow of funding impact the structure of a cross-cultural survey. Below are examples of how five large-scale, cross-cultural survey programs have been funded. Please see the websites of these programs for further information.

- The European Social Survey [21] investigates the interaction between Europe's changing institutions and the attitudes, beliefs, and behavior patterns of its diverse populations using face-to-face interviews in over 30 countries throughout four rounds. Funding for the central coordinating center has come from the European Commission's Framework Programs and the European Science Foundation. National scientific bodies have funded their own country's data collection and coordination.

- The International Social Survey Programme (ISSP) [22] investigates current social science topics in each of 43 participating countries by collecting self-administered questionnaires. Each survey organization has funded all of its own costs. There are no central funds.

- Latinobarómetro [23] investigates social development with face-to-face interviews in 18 Latin American countries occurring sporadically. Initial funding came from the European Commission. There have been several additional funding sources, including: international organizations (e.g., Inter-American Development Bank, United Nations Development Programme, World Bank), government agencies, and private sector sources.

- The Survey of Health, Ageing, and Retirement in Europe [24] investigates respondents in an aging population (50 and over) in 11 countries throughout three waves (2004, 2006-2007, and 2008-2009). The European Union has funded the data collection under the European Commission and funded the analyses under Advanced Multidisciplinary Analysis of New Data on Ageing. The U.S. National Institute on Aging has provided additional funding; other national funding agencies provided support as well.

- The World Mental Health Surveys [25] investigate mental disorders with face-to-face interviews in 28 countries since 2000. Funding for the data collection and analysis coordinating centers has come from the World Health Organization. Several additional funding sources have included the U.S. National Institute of Mental Health, European Commission, MacArthur Foundation, Robert Wood Johnson

Foundation, World Health Organization, Pan American Health Organization, various pharmaceutical companies, and governments of the participating countries. Each participating country has had its own source of funding.

Appendix C

Organizational structures

Below are descriptions of the organizational structures that have been used on three large-scale, cross-cultural survey programs. These examples are only illustrative. Please visit the survey programs' websites for more information about their structure.

- European Social Survey [21]
  - The Central Coordinating Team is responsible for overseeing the entire study. The Central Coordinating Team is in contact with the Funders, the Scientific Advisory Board, the Specialist Advisory Groups, and the National Coordinators/Survey Institutes.
  - The Scientific Advisory Board consists of representatives from each participating country, two representatives from the European Commission, and two representatives from the European Science Foundation.
  - The Specialist Advisory Groups has separate teams with expertise in question design, methods, sampling, and translation.
  - The National Coordinators/Survey Institutes have one director from each country and one national survey organization from each country. The survey organizations are chosen by their country's respective national academic funding body.

- Survey of Health, Ageing, and Retirement in Europe [3] [24]
  - The Coordinating Center oversees the entire study and reconciles differences between Country Teams and Cross-national Working Groups. The Coordinating Center is led by the Co-ordinator. Members of the group are internationally-recognized experts in their fields. The Co-ordinator receives additional support from CentERdata, the Survey Research Center, the Centre for Survey Research and Methodology, and the National Centre for Social Research.
  - Country Teams and Cross-national Working Groups form a matrix organizational structure. Country Teams are led by Country Team Leaders. They are responsible for carrying out the study in their respective country and select one national survey organization to conduct the survey.
  - Cross-national Working Groups are led by Working Group Leaders. There is a working group for each topic covered in the questionnaire, and each respective working group is responsible for their topic's module. Additionally, there are working groups for methodological concerns. The Cross-National Working Groups are set up so each country can have a topic-specialist in each working

group, but it is not always the case that each country has expert in that field.
- The Advisory Panels are available if guidance is needed from those with experience in a given area. There are Advisory Panels with representatives from Survey Methodology and Quality Control, as well as from the Health and Retirement Study, the English Longitudinal Survey on Ageing, and the respective Countries.

- World Mental Health Surveys [17] [25]
  - The World Health Organization is invested in the objectives of this survey and works closely with two study-level Principal Investigators. These study-level researchers make many of the ultimate decisions for the entire study. The World Health Organization is in contact with the Data Collection Coordination Center and the Analysis Coordination Center.
  - The Data Collection Coordination Center is instrumental in writing and implementing the specifications for pre-production and production activities. The University of Michigan is the Data Collection Coordination Center and its tasks include such activities as selecting survey organizations, training interviewers, and providing assistance during data collection.
  - The Analysis Coordination Center makes decisions regarding post-production activities. Harvard University is the Analysis Coordination Center.
  - The Working Groups are analysis teams that focus on one particular aspect or analytic perspective of mental health. Each Working Group is led by a Chair. Examples of focal topics include the following: ADHD, drug dependence, gender, social class, suicide, and personality disorders. The Working Groups are in contact with the Analysis Coordination Center and the Principal Investigators from each country.
  - The Principal Investigators from each country oversee their respective country's survey.
  - The Data Collection Organizations are the survey organizations within each country that carry out the field operations.

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Audit trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Coding	Translating nonnumeric data into numeric fields.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Complex survey data (or designs)	Survey datasets (or designs) based on stratified single or multistage samples with survey weights designed to compensate for unequal probabilities of selection or nonresponse.

Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Cooperation rate	The proportion of all elements interviewed of all eligible units ever contacted.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Item nonresponse, item missing data	The lack of information on individual data items for a sample element where other data items were successfully obtained.

Interface design	Aspects of computer-assisted survey design focused on the interviewer's or respondent's experience and interaction with the computer and instrument.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Mode	Method of data collection.
Noncontact	Sampling units that were potentially eligible but could not be reached.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Outcome rate	A rate calculated based on the study's defined final disposition codes that reflect the outcome of specific contact attempts before the unit was finalized. Examples include response rates (the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.), cooperation rates (the proportion of all units interviewed of all eligible units ever contacted), refusal rates (the proportion of all units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units), and contact rates (the proportion of all units are reached by the survey).
Overrun	The exceeding of costs estimated in a contract.
Paradata	Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.
Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins.

Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Refusal rate	The proportion of all units of all potentially eligible sampling units in which a respondent sampling unit refuses to do an interview or breaks off interviews of all potentially eligible sampling units.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.

Restricted-use data file	A file that includes information that can be related to specific individuals and is confidential and/or protected by law. Restricted-use data files are not required to include variables that have undergone coarsening disclosure risk edits. These files are available to researchers under controlled conditions.
Sample design	Information on the target and final sample sizes, strata definitions and the sample selection methodology.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.

Stratification	A sampling procedure that divides the sampling frame into mutually exclusive and exhaustive groups (or strata) and places each element on the frame into one of the groups. Independent selections are then made from each stratum, one by one, to ensure representation of each subgroup on the frame in the sample.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Survey weight	A statistical adjustment created to compensate for complex survey designs with features including, but not limited to, unequal likelihoods of selection, differences in response rates across key subgroups, and deviations from distributions on critical variables found in the target population from external sources, such as a national Census.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Task	An activity or group of related activities that is part of a survey process, likely defined within a structured plan, and attempted within a specified period of time.
Tender	A formal offer specifying jobs within prescribed time and budget.
Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.

Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Working group	Experts working together to oversee the implementation of a particular aspect of the survey lifecycle (e.g., sampling, questionnaire design, training, quality control, etc.)

References

[1] Aitken, A., Hörngren, J., Jones, N., Lewis, D., & Zilhäo, M. J. (2003). *Handbook on improving quality by analysis of process variables.* Luxembourg: Eurostat. Retrieved March 27, 2010, from http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20ON%20IMPROVING%20QUALITY.pdf

[2] Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.

[3] Börsch-Supan, A., Jürges, H., & Lipps, O. (2003). *SHARE: Building a panel survey on health, aging and retirement in Europe*. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging. Retrieved March 26, 2010, from http://www.mea.uni-mannheim.de/mea_neu/pages/files/nopage_pubs/k3pzhnwk1t1zjymk_dp32.pdf

[4] Cleland, D. I., & Gareis, R. (Eds.). (2006). *Global project management handbook* (2nd ed.). New York, NY: McGraw-Hill.

[5] Deming, E. (1986). *Out of the crisis*. Cambridge, MA: Cambridge University Press.

[6] Federal Committee on Statistical Methodology. (1983). *Statistical policy working paper 9: Contracting for surveys*. Washington, DC: Office of Management and Budget.

[7] Fitzgerald, R., & Jowell, R. *Measurement equivalence in comparative surveys: The European Social Survey (ESS)—from design to implementation and beyond.* Unpublished manuscript.

[8] Griffin, R.W. & Moorehead, G. (2009). *Organizational behavior – Managing people and organizations* (8th ed.). Cincinnati, OH: South-Western Educational Publishing.

[9] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[10] International Organization for Standardization. (2006). *Market, opinion and social research—terms, definitions and service requirements* (ISO 20252). Geneva. Retrieved April 30, 2010, from https://www.astandis.at/shopV5/Preview.action%3Bjsessionid=0D8F503B3BD48375403D96257760FABB?preview=&dokkey=342276&selectedLocale=en/

[11] Jowell, R. (1998). How comparative is comparative research? *American Behavioral Scientist, 42*(2), 168-177.

[12] Defeo, J. A. & Juran , J. M. (2010). *Juran's quality handbook: The complete guide to performance excellence* (6th ed.). New York, NY: McGraw-Hill.

[13] Lyberg, L. E., & Biemer, P. P. (2008). Quality assurance and quality control in surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology*. New York/London: Lawrence Erlbaum Associates.

[14] Mohler, P. Ph., Pennell, B-E., & Hubbard, F. (2008). Survey documentation: Toward professional knowledge management in sample surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology.* New York/London: Lawrence Erlbaum Associates.

[15] Montgomery, D.C. (2005). Introduction to statistical quality control (5th ed.). Hoboken, NJ: John Wiley & Sons.

[16] Office of Management and Budget. (2006). *Standards and guidelines for statistical surveys*. Washington, DC: Office of Information and Regulatory Affairs, OMB. Retrieved January 10, 2010, from http://www.whitehouse.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf

[17] Pennell, B-E., Mneimneh, Z., Bowers, A., Chardoul, S., Wells, J. E., Viana, M. C., et al. (2009). Implementation of the World Mental Health survey initiative. In R. C. Kessler & T. B. Üstün (Eds.), *Volume 1: Patterns of mental illness in the WMH surveys.* Cambridge, MA: Cambridge University Press.

[18] Ryan, T.P. (2000). *Statistical methods for quality improvement* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[19] United Nations. (2005). *Household surveys in developing and transition countries.* New York, NY: United Nations, Department of Economic and Social Affairs. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[20] Worcester, R., Lagos, M., & Basañez, M. (2000). *Cross-national studies: Lessons learned the hard way*. Paper presented at the 2000 WAPOR/AAPOR Conference, Portland, OR.

Internet Links

[21] European Social Survey (ESS). (2010). Retrieved May 03, 2010, from http://www.europeansocialsurvey.org

[22] International Social Survey Programme (ISSP). (2010). Retrieved May 04, 2010, from http://www.issp.org

[23] Latinobarómetro. (2010). Retrieved May 4, 2010, from http://www.latinobarometro.org

[24] Survey of Health, Ageing, and Retirement in Europe. (2010). Retrieved May 4, 2020, from http://www.share-project.org

[25] World Mental Health Study. (2010). Retrieved May 04, 2010, from http://www.hcp.med.harvard.edu/wmh/

Further Reading

Heath, A., Fisher, S., & Smith, S. (2005). The globalization of public opinion research. *Annual Review of Political Science, 8*, 297-333.

Lynn, P. (2003) Developing quality standards for cross-national survey research: five approaches. *International Journal of Social Research Methodology, 6*(4), 323-336. Retrieved October 8, 2009, from http://www.soc.uoc.gr/socmedia/papageo/developing%20quality%20standards%20for%20cross-national%20surveys.pdf

II. Survey Quality

Sue Ellen Hansen, Grant Benson, Ashley Bowers, Beth-Ellen Pennell, Yuchieh Lin, and Benjamin Duffey

Introduction

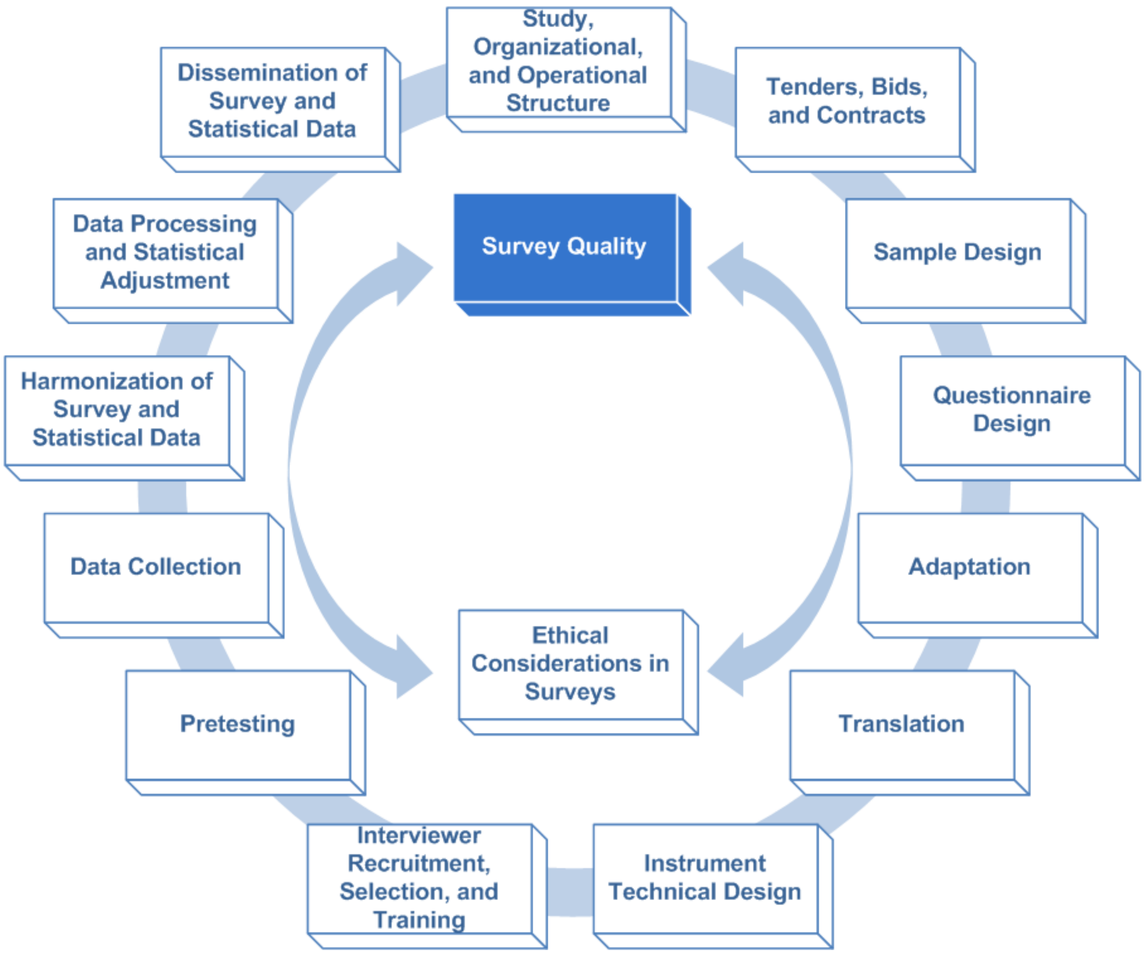
This chapter presents a quality framework for assessing quality in cross-cultural surveys, followed by guidelines for managing and assessing quality throughout the survey lifecycle.

In mono-cultural surveys, assessing the quality of survey data requires adequate documentation of the entire survey lifecycle and an understanding of protocols used to assure quality. In such surveys, there may be challenges to overcoming methodological, organizational, and operational barriers to ensuring quality. For example, a country may not have the infrastructure or an organization may not have the means to implement a study entirely according to survey best practices.

In cross-cultural survey research, the challenges increase. Cross-cultural surveys hinge on the comparability or equivalence of data across cultures. Moreover, cross-cultural survey quality assessment procedures and criteria become more complex with additional survey processes, such as adaptation and translation of questions and harmonization of data across multiple surveys (see Adaptation, Translation, and Data Harmonization).

Figure 1 shows the survey production lifecycle as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose [20]), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications.

Figure 1. The Survey Lifecycle



Quality Framework

The framework adopted by these guidelines for assuring and assessing quality is informed by research on survey errors and costs and quality management, and highlights three aspects of quality: total survey error ([14] [15]), fitness for intended use ([9]; also known as "fitness for purpose" [20]), and survey process quality ([4] [19] [23]).

Total survey error

The total survey error (TSE) paradigm is widely accepted as a conceptual framework for evaluating survey data quality [2] [6]. TSE defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest, which is the sum of random errors (variance) and squared systematic errors (bias). TSE takes into consideration both measurement (construct validity, measurement error, and processing error)—i.e., how well survey questions measure the constructs of interest—and representation (coverage error, sampling error, nonresponse error, and adjustment error) [15]—i.e., whether one can generalize to the target population using sample survey data. In the TSE perspective, there may be cost-error tradeoffs, that is, there may be tension between reducing these errors and the cost of reducing them.

With advances in computerized interviewing software and sample management systems, data related to quality increasingly can be collected with survey data, and can be used to measure various components of error. These include paradata [4] [5], data from experiments embedded in a survey, and supplementary data, such as nonresponse followup questions. Each of these facilitates evaluation of survey data in terms of TSE.

Fitness for intended use

Biemer and Lyberg [4] argue that the TSE framework lacks a user perspective, and that it should be supplemented by using a more modern quality paradigm—one that is multidimensional and focuses on criteria for assessing quality in terms of the degree to which survey data meet user requirements (fitness for intended use). By focusing on fitness for intended use, study design strives to meet user requirements in terms of survey data accuracy and other dimensions of quality (such as comparability and timeliness). In this perspective, ensuring quality on one dimension (comparability) may conflict with ensuring quality on another dimension (timeliness); and there may be tension between meeting user requirements and the associated cost of doing so on one or more dimensions. There are a number of multidimensional quality frameworks in use across the world (see, for example, [5] [7] [16] [27] [28]).

Table 1 shows seven dimensions that are often used to assess the quality of national official statistics in terms of both survey error and fitness for use: comparability, relevance, accuracy, timeliness and punctuality, accessibility, interpretability, and coherence. In this framework, TSE may be viewed as being covered by the accuracy dimension.

Table 1. Dimensions of Quality

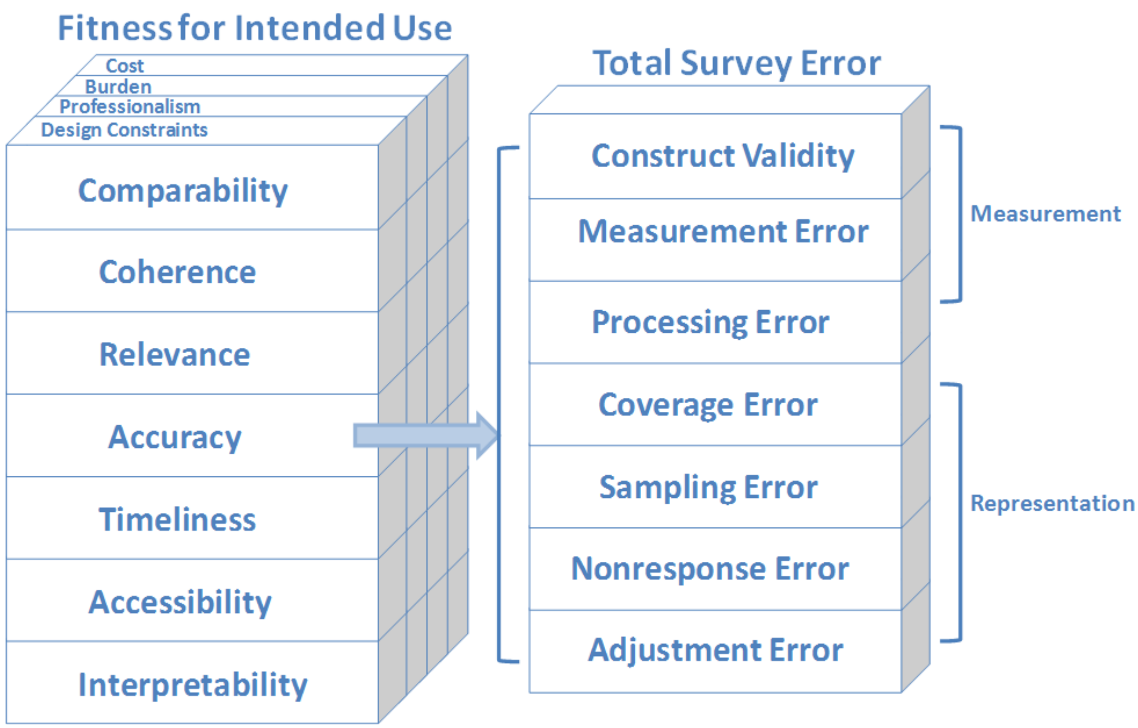
Quality Dimension	Description
Comparability	Are the data from different countries or cultures comparable to each other (equivalent)?
Coherence	Do the data form a coherent body of information that can be rearranged or combined with other data?
Relevance	Do the data meet the requirements of the client and users?
Accuracy	Are the data describing the phenomena that they were designed to measure; that is, are the survey estimates close to the true values of the population parameters they are meant to measure?
Timeliness and punctuality	How much time has elapsed between the end of the data collection and when the data are available for analysis? Are the data available when expected, based on client specifications?
Accessibility	Can users easily obtain and analyze the data?
Interpretability	Do the data make sense in terms of users' hypotheses? Are supplementary data available to facilitate analysis, e.g., data that describe the major characteristics and structure of the data (metadata) as well as data about the survey processes (paradata)?

Cost, burden, professionalism, and design constraints are factors that may also affect fitness for use on these dimensions:

- **Cost** – are monetary resources optimized?
- **Burden** – are interviewer and respondent burden minimized?
- **Professionalism** – are staff provided with clear behavioral guidelines and professional training, are there adequate provisions to ensure compliance with relevant laws, and is there demonstration that analyses and reporting have been impartial?
- **Design Constraints** – are there context-specific constraints on survey design that may have had an impact on quality (for example, using a different mode of interview in one culture than in others)?

The aim is to optimize costs, minimize burden and design constraints where appropriate—based on the need to be sensitive to local survey contexts, and to maximize professionalism. Figure 2 shows the dimensions of quality and factors that affect quality in terms of fitness for use (see [3] [5] [7] [16] [27] [28] for examples of dimensions of quality used by statistical agencies). It also shows the accuracy dimension in terms of TSE [2] [14] [15].

Figure 2. Fitness for Intended Use (Quality Dimensions) and Total Survey Error (Accuracy Dimension)



The dimensions of quality (comparability, coherence, relevance, accuracy, and so on) and factors that may have an impact on quality (cost, burden, professionalism, and design constraints) apply to all surveys. However, in a cross-cultural context, challenges increase:

- The quality dimensions of **coherence** and **comparability** are the *raison d'être* for cross-national and cross-cultural survey research. Fitness for intended use cannot be met without quality on these dimensions.
- **Relevance** may be harder to achieve in comparative research, in that decisions have to be made about what level of relevance to aim for with a standardized survey across many cultures and countries.
- **Accuracy** in terms of TSE may be difficult to estimate consistently across cross-cultural surveys.

- **Timeliness and punctuality** may be a challenge in cross-national research; for example, data collection may occur in vastly different climates or with varying organizational infrastructures (see Data Collection).

- **Accessibility** in the cross-national context can mean more than simply making survey data publicly available, particularly in majority countries, where it also may be necessary to include capacity building or data user training to make the data truly accessible to local users. Country-level data access laws and regulations may also come into play (see Data Dissemination).

- **Interpretability** of data may be difficult without metadata documentation about the data that would facilitate comparison across cross-cultural surveys (see Data Dissemination).

Appendix A highlights recommendations from specific chapters in these guidelines in relation to dimensions of quality.

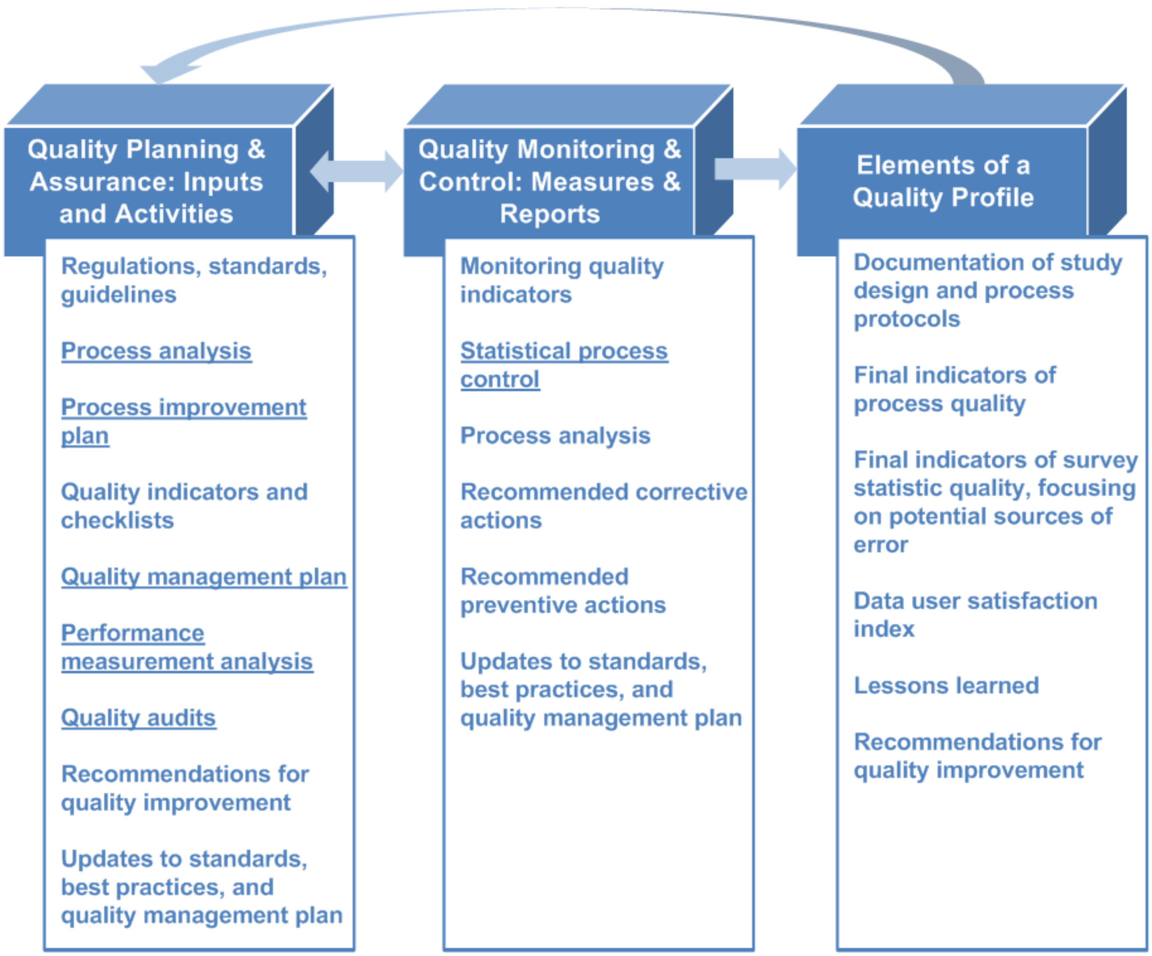
Survey process quality

Fitness for intended use provides a general framework for assessing the quality of cross-cultural surveys, and defines the essential dimensions of quality, one of which is accuracy (TSE). A third approach to quality monitoring and assessment is survey process quality management, and the notion of continuous process improvement ([15]).This approach focuses on quality at three levels: the organization, the process, and the product [18]. Quality products cannot be produced without quality processes, and having quality processes requires an organization that manages for quality.

A focus on survey production process quality requires the use of quality standards and collection of standardized study metadata, question metadata, and process paradata [7]. Figure 3 shows the elements of survey process quality management that allow users to assess the quality of processes throughout the survey lifecycle: quality assurance, quality control [17] [18], and a quality profile [4] [11].These are discussed further in the guidelines below.

Cross-cultural survey organizations may vary in what cost-quality tradeoffs they can make, as well as processes they generally monitor for quality purposes. However, if each organization reaches a minimum standard through adherence to the quality guidelines of the study's coordinating center, the coordinating center can assess the quality of each survey based on quality indicators (paradata) from each organization, and create a quality profile that allows users to assess survey data quality and comparability across cultures. Appendix B summarizes for each chapter examples of elements of quality planning and assurance, quality monitoring and control, and a quality profile.

Figure 3. Survey Process Quality Management



Guidelines

Goal: To ensure the quality of survey production processes and consequently the survey data throughout the survey lifecycle, as well as clear and comprehensive documentation of study methodology, and to provide indicators of process and data quality.

1. Develop a sustainable quality management plan.

Rationale

Developing planned, systematic quality assurance (Guideline 2) and quality control (Guideline 3) activities helps ensure that the study and survey data meet client and user requirements. It also facilitates development of a quality profile (Guideline 4), which should document survey methodology, key indicators of quality, lessons learned, and recommendations for improvement.

Procedural Steps

- Review available cross-cultural survey standards and best practices for ensuring the quality of survey processes, survey data, and documentation (such as these guidelines).
- Review existing quality profiles (Guideline 4) and lessons learned from other studies. Use standardized quality profiles and protocols to establish sustainable quality management.
- Review study requirements for quality assurance and quality control. These may be developed at the study design stage by the coordinating center, the survey organization, or both.
- Review study goals and objectives, required products and deliverables, and study timeline and budget.
- Review country-specific regulations and legislation relevant to conducting survey research.
- Through analysis of the processes in the survey lifecycle (process analysis) [1], identify characteristics of survey products (e.g., coded data) that could vary during the processes (e.g., verification failures). For example,
    - Use tools to analyze a process, to determine what steps in the process need to be monitored to ensure quality, and to identify quality indicators to monitor [1]. Examples of tools used to analyze processes are:

- Cause and effect diagrams ("fishbone" diagrams).
- Flow charts.

- Identify key indicators of the quality of the product(s) of the process, in terms of TSE and other dimensions of quality, as well as factors such as cost, burden, and the risk of not meeting quality requirements. See Appendix A for examples of survey quality indicators as they relate to TSE and the fitness for use quality dimensions (see Quality Framework).
- If possible, use such indicators to determine whether the process is stable or in control; that is, is variation on a key indicator due to randomness alone? This can be done using paradata from similar studies the organization has conducted or is conducting, or from pilot studies.
- Define measurement and reporting requirements for use during quality assurance (see Guideline 2) and quality control (see Guideline 3), and determine who would be responsible for ensuring that quality assurance and quality control activities are carried out.
- Assess whether these requirements can be met through current procedures and systems, and with currently collected paradata; and if not, develop a process improvement plan.
  - Create cost/error tradeoff decision rules about how to alter the features of the study design if the goals are not met.
- Use quality planning tools to help determine what performance analyses and assessments should be used. For example,
  - A cost-benefit analysis of potential quality management procedures and activities; that is, evaluating their benefits in relation to the cost of performing them relative to overall study costs.
  - Benchmarking, that is, comparing planned activities against those of similar studies, and the outcomes of those activities, to form a basis for performance measurement.
  - Statistical analysis of factors that may influence indicators of process or product quality.
  - Cost of quality and cost of poor quality analyses.

- Develop a quality assurance plan, which could include (see Appendix B):
  - The process improvement plan.
  - Performance and product quality baselines.
  - Process checklists.
  - A training plan.
  - Recommended performance analyses and assessments (e.g., quality assurance procedures for verifying interviews and evaluating interviewer performance).
  - Required process quality audits, reviews, and inspections (e.g., review of tapes of interviews to assess interviewer performance).

- Develop a plan for continuous monitoring of processes to ensure that they are stable and that products are meeting requirements (Quality Control; see [1], Guideline 3, and Appendix B). Such a plan could include:
  - The process improvement plan.
  - Performance and product quality baselines.
  - Quality indicators identified in process analysis and planning for responsive design.
  - Performance analyses and assessments to use to monitor processes.
  - Tools to use to monitor processes and product quality, e.g., Pareto charts and statistical process control charts.
  - Reports to prepare on performance measurement, such as interviewer training certification.

- Develop procedures to ensure that throughout the survey lifecycle all documentation, reports, and files related to quality planning and assurance, quality monitoring and control, and process improvement are retained. This facilitates preparing a quality profile for users of the disseminated survey data (see Guideline 4 and Data Dissemination).

- Develop procedures for updating the quality management plan as needed during the survey lifecycle.

Lessons Learned

- There are many quality management methodologies that survey organizations may use that focus on the three levels of quality: product, process, and organization; for example, Total Quality Management (TQM). Discussion of such methodologies is beyond the scope of this chapter, but experience has shown that they can help organizations manage for quality.

- Developing a quality management plan alone does not necessarily guarantee quality. Other project management practices may also affect quality. Many survey organizations and statistical agencies have recognized the value of also adhering to professional project management guidelines, such as those of the Project Management Institute (PMI) [26] and the International Project Management Association (IPMA). Many have certified project managers and follow professional project management best practices that may affect quality, schedule, and costs, such as developing risk management and communication plans. As with a quality management plan, these can be critical to ensuring the quality of processes and survey data.

2. Perform quality assurance activities.

Rationale

Quality assurance is the planned procedures and activities (see Guideline 1) an organization uses to ensure that the study meets process and product quality requirements. It specifies ways in which quality can be measured.

Procedural Steps

- For each process in the survey lifecycle, perform quality assurance activities as outlined in the quality management plan, such as (see Appendix B):
  - Certification by the coordinating center that an organization's study design and quality standards meet study standards (see Study, Organizational, and Operational Structure).
  - Pretest consent protocols and forms to ensure comprehension (see Ethical Considerations in Surveys).

- Perform performance and product quality assessments. Examples are:
  - Certification of interviewers after training (rate of certification, rate of certification after follow-up training, etc.); that is, based on evaluation of interviews (taped or monitored), determination that the interviewer is ready to work on the study.
  - Verification of coded questionnaires (rate of verification failures).

- Generate indicators of quality for each assessment, based on baselines established in quality planning (Guideline 1), and create reports on performance and quality assessments, which can be used for both quality monitoring and control (see Guideline 3), and documentation in a quality profile (see Guideline 4).

  Perform quality audits at key points in the survey lifecycle if study guidelines for quality management require them. These generally are structured independent reviews to determine whether activities comply with study and organizational policies and procedures for managing quality. They are intended to identify inefficiencies in processes, and to make recommendations for reducing the cost of quality management and increasing the quality of processes and products. In international studies, these generally would be done by the survey organization, or an independent local auditor.

- Provide documentation for:
  - Performance and quality assessments.

- Recommended corrective actions and corrective actions taken.
- Updates to baselines.
- Changes to quality assurance plan.

3. Perform quality control activities.

Rationale

To ensure that standards and requirements are met, it is necessary to monitor study processes and the products produced against predetermined baselines and requirements, and continuously evaluate whether processes are stable (in control) and quality requirements are being met [4] [17]. This may lead to recommendations for preventing or minimizing error or inefficiencies, updates to the quality management plan (see Guideline 1), and suggestions for improving standards and best practices. The result is continuous process improvement ([4] [17] [23]), through improved quality assurance (see Guideline 2) and improved quality monitoring and control.

As indicated in Figure 3, quality control is closely linked to quality assurance, and the outputs of each feed into the other. Thus, in some respects, quality control may be viewed as part of quality assurance. However, these are separated in this chapter to make monitoring and controlling performance and product quality an explicit part of quality management.

Procedural Steps

- Perform quality monitoring and control activities as outlined in the quality management plan, such as (see Appendix A for examples):
  - Monitor process quality indicators (see Guideline 1).
  - Analyze and report on results of quality assurance activities, such as interviewer training certification, data entry verification, checking that a process met specifications, etc.
  - In accordance with the quality management plan (see Guideline 1), generate charts and graphs to monitor processes. Examples of such tools are [1]:
    - Pareto chart
    - Statistical process control chart

- Perform process analysis (see Guideline 2) if quality requirements are not being met.

- Determine whether there is a need to:
  - Recommend corrective actions.
  - Modify the process improvement plan.

- Modify the quality management plan.

- Provide documentation for:
  - Performance and quality assessments.
  - Recommended corrective actions and corrective actions taken.
  - Updates to baselines.
  - Changes to the quality management and quality assurance plans.

Lessons learned

- Some organizations have used quality control techniques to monitor survey data collection processes and adapt study designs when quality goals are not met. This is known as adaptive or responsive survey design [13].

4. Create a quality profile

Rationale

A quality profile (also known as a quality report) synthesizes information from other sources, documenting survey methodology used throughout the survey lifecycle, providing indicators of process and data quality (sampling and nonsampling errors), corrective actions taken, lessons learned, and recommendations for improvement and further research. It provides the user all information available to help assess data quality in terms of fitness for intended use, total survey error, and other factors (see Framework above). See [9] for an example of guidelines for such reports, [10], [11], and [29] for examples of quality profiles, and Appendix A for examples from chapters in these guidelines.

Procedural Steps

- Document procedures and methodology used for key stages or processes in the lifecycle (see Appendix B). For example, for sample design this would include:
  - Time dimension of design (e.g., one time cross sectional, fixed or rotating panel)
  - Target and survey population definitions, including inclusion/exclusion criteria.
  - Sampling frame(s) descriptions.
  - Maps and protocol used in field listing.
  - Description of all stages of selection, including sample sizes, stratification, clustering, oversampling and number of replicates fielded at each stage.

- Documentation of procedures to determine probabilities of selection and weights for each stage of selection.
- Tables of the precision of the estimates of key survey statistics.
- (If necessary), descriptions of substitution procedures.

For each process documented, this should include

- Quality assurance procedures.
- Quality control procedures.
- Corrective actions taken.

- Provide key indicators of quality for all dimensions of quality (see [9] and Appendix B), some of which can be collected during data collections, others afterwards. They include:
  - Comparability.
  - Coherence.
  - Relevance.
  - Accuracy (see Quality Framework), including
    - Measurement error
    - Processing error
    - Coverage error
    - Sampling error
    - Nonresponse error
    - Adjustment error
  - Timeliness and punctuality.
  - Accessibility.
  - Interpretability.

- Document lessons learned and make recommendations for improvement in studies of the same design, and, if possible, make recommendations for methodological research that could inform design of similar studies in the future. Such information would be useful for the study's coordinating center and national survey agencies, but also researchers and organizations interested in conducting similar studies.

Appendix A

The following table lists recommendations from individual chapters in these guidelines that are related to the dimensions of quality. Also included are examples of indicators of quality adapted from Eurostat's standard quality indicators [12].

Quality Dimension	Guidelines
Comparability To ensure as much as possible, that observed data from different countries or cultures are comparable (equivalent). Indicators: Time ▪ The differences, if any, in concepts and methods of measurements between last and previous reference period ▪ A description of the differences, including an assessment of their effect on the estimates Geographical ▪ All differences between local practices and national standards (if such standards exist) ▪ An assessment of the effect of each reported difference on the estimates Domains ▪ A description of the differences in concepts and methods across cross-cultural surveys (e.g., in classifications, statistical methodology, statistical population, methods of data manipulation, etc.) ▪ An assessment of the magnitude of the effect of each difference	Establish minimum criteria for inclusion in a cross-national survey dataset, if applicable, as follows: Minimize the amount of undue intrusion by ensuring comparable standards when appropriate (based on differences in local survey contexts) for informed consent and resistance aversion effort, as well as other potentially coercive measures such as large respondent incentives (see Ethical Considerations in Surveys). Define comparable target populations and verify that the sampling frames provide adequate coverage to enable the desired level of generalization (see Sample Design). Minimize the amount of measurement error attributable to survey instrument design, including error resulting from context effects, as much as possible (see Instrument Technical Design). Minimize or account for the impact of language differences resulting from potential translations (see Translation and Adaptation). Minimize the effect interviewer attributes have on the data through appropriate recruitment, selection, and case assignment; minimize the effect that interviewer behavior has on the data through formal training (see Interviewer Recruitment, Selection, and Training). Identify potential sources of unexpected error by implementing pretests of translated instruments or instruments fielded in different cultural contexts (see Pretesting). Reduce the error associated with nonresponse as much as possible (see Data Collection for a discussion of nonresponse bias and methods for increasing response rates). Minimize the effect that coder error has on the data through appropriate coder training (see Data Processing and Statistical Adjustment).

Quality Dimension	Guidelines
(Comparability)	If possible, provide a crosswalk between survey instruments fielded at different times or for different purposes, but using the same questions, to facilitate analysis and post-survey quality review (see Data Dissemination).
Coherence To ensure that the data can be combined with other statistical information for various, secondary purposes. Indicators: ▪ A description of every pair of statistics (statistical unit, indicator, domain, and breakdown) for the survey(s) that should be coherent ▪ A description of any of the differences that are not fully explained by the accuracy component. ▪ A description of the reported lack of coherence, for specific statistics	Create a clear, concise description of all survey implementation procedures to assist secondary users. The Study, Organizational, and Operational Structure chapter lists topics which should be included in the study documentation; there are also documentation guidelines within each chapter. Provide data files in all the major statistical software packages and test all thoroughly before they are made available for dissemination (see Sample Design, and Data Dissemination). Designate resources to provide user support and training for secondary researchers (see Data Dissemination). See Data Harmonization for a discussion of the creation of common measures of key economic, political, social, and health indicators.
Relevance To ensure that the data meet the needs of the client or users. Indicators: ▪ A description of clients and users ▪ A description of users' needs (by main groups of users) ▪ An assessment of user satisfaction	Clearly state the study's goals and objectives (see Study, Organizational, and Operational Structure). Conduct a competitive bidding process to select the most qualified survey organization within each country or location (see Tenders, Bids, and Contracts). While designing the questionnaire, ensure all survey questions are relevant to the study objectives (see Questionnaire Design). Construct the data file with a data dictionary of all variables in the selected element data file, with all variable names and an accompanying description which are relevant to the study objectives (see Sample Design).
Accuracy To ensure that the data describe the phenomena they were designed to measure. This can be assessed in terms of Mean Square Error (MSE).	Pretest all the versions of the survey instrument to ensure that they adequately convey the intended research questions and measure the intended attitudes, values, reported facts and/or behaviors (see Pretesting).

Quality Dimension	Guidelines
(Accuracy) Indicators: Measurement error: ▪ A description of the methods used to assess measurement errors (any field tests, reinterviews, split sample experiments, or cognitive laboratory results, etc.) ▪ A description of the methods used to reduce measurement errors ▪ Average interview duration ▪ An assessment of the effect of measurement errors on accuracy Processing Error: ▪ A description of the methods used to reduce processing errors ▪ A description of the editing system ▪ The rate of failed edits for specific variables. ▪ The error rate of data entry for specific variables and a description of estimation methodology ▪ The error rate of coding for specific variables and a description of the methodology followed for their estimation ▪ A description of confidentiality rules and the amount of data affected by confidentiality treatment Coverage error: ▪ A description of the sampling frame ▪ Rates of over-coverage, under-coverage, and misclassification broken down according to the sampling stratification ▪ A description of the main misclassification and under- and over-coverage problems encountered in collecting the data ▪ A description of the methods used to process the coverage deficiencies Sampling error: ▪ Type of sample design (stratified, clustered, etc.) ▪ Sampling unit at each stage of sampling ▪ Stratification and sub-stratification criteria ▪ Selection schemes ▪ Sample distribution over time ▪ The effective sample size ▪ Coefficients of variation of estimates and a description of the method used to compute them (including software) ▪ An assessment of resulting bias due to the estimation method	In order to reliably project from the sample to the larger population with known levels of certainty/precision, use probability sampling (see Sample Design). Provide a report on each variable in the dataset of selected elements to check correct overall sample size and within stratum sample size, distribution of the sample elements by other specific groups such as census enumeration areas, extreme values, nonsensical values, and missing data (see Sample Design). If possible, assess accuracy by looking at the differences between the study estimates and any available "true" or gold standard values (see Data Collection).

Quality Dimension	Guidelines
(Accuracy) Nonresponse error: ▪ Unit nonresponse rate ▪ Identification and description of the main reasons for nonresponse (e.g., non-contact, refusal, unable to respond, non-eligible, other nonresponse) ▪ A description of the methods used for minimising nonresponse ▪ Item nonresponse rates for variables ▪ A description of the methods used for imputation and/or weighting for nonresponse ▪ Variance change due to imputation ▪ An assessment of resulting bias due to nonresponse Model assumptions error: ▪ A description of the models used in the production of the survey's statistics ▪ A description of assumptions used on which the model relies ▪ A description of any remaining (unaccounted for) bias and variability which could affect the statistics	
Timeliness and punctuality To ensure that the data are available for analysis when they are needed. Indicators: ▪ The legal deadline imposed on respondents ▪ The date the questionnaires were sent out ▪ Starting and finishing dates of fieldwork ▪ Dates of processing ▪ Dates of quality checks ▪ The dates the advance and detailed results were calculated and disseminated ▪ If data is transmitted later than required by regulation or contract, the average delay in days or months in the transmission of results with reference to the legal deadline ▪ If data are transmitted later than required by regulation or contract, the reasons for the late delivery and actions taken or planned for the improving timeliness	Time data collection activities appropriately (see Data Collection, and Pretesting). Create a study timeline, production milestones, and deliverables with due dates (see Study, Organizational, and Operational Structure).

Quality Dimension	Guidelines
Accessibility To ensure that the data can easily be obtained and analyzed by users. Indicators: ▪ A description of how to locate any publication(s) based on analysis of the data ▪ Information on what results are sent to reporting units included in the survey ▪ Information on the dissemination scheme for the results ▪ A list of variables required but not available for reporting ▪ Reasons why variables are not available	Save all data files and computer syntax from the preferred statistical software package needed during sample design process in safe and well labeled folders for future reference and use (see Sample Design). Establish procedures early in the survey lifecycle to insure that all important files are preserved (see Data Dissemination). Test archived files periodically to verify user accessibility (see Data Dissemination). Create electronic versions of all project materials whenever feasible (see Data Dissemination). Produce and implement procedures to distribute restricted-use files, if applicable (see Data Dissemination).
Interpretability To ensure that supplementary metadata and paradata are available to analysts. Indicator: • A copy of any methodological documents relating to the statistics provided	At the data processing stage of the study, create a codebook that provides question-level metadata matched to variables in the dataset. Metadata include variable names, labels, and data types, as well as basic study documentation, question text, universes (the characteristics of respondents who were asked the question), the number of respondents who answered the question, and response frequencies or statistics (see Sample Design and Data Processing and Statistical Adjustment). Collect and make available process data collected during data collection, such as timestamps, keystrokes, and mouse actions ("paradata") (see Instrument Technical Design).

Appendix B

The following table summarizes recommended elements of process quality management relevant to each chapter in these guidelines. These are meant to reflect quality management at two levels: (1) the overall study level; and (2) the national organization level. It is not meant to convey that all elements listed should be part of a study's design, but to provide examples and to help guide the development of specifications for quality management for a study.

If possible, the study's quality profile (quality report) would include a summary of each organization's performance, based on standardized quality indicators. It also would include lessons learned and recommendations for improvement.

Where possible, examples are taken from the individual chapters in these guidelines. Not all chapters have specific measures for monitoring and controlling quality. Even without clear individual rates or measures of quality, there often may be reports on quality assurance activities that facilitate assessing quality.

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
Study, Organizational, and Operational Structure	**Inputs** ▪ Study goals and objectives ▪ Country-specific legislation on conducting survey research ▪ Leadership, roles, and responsibilities ▪ Timeline ▪ Deliverables ▪ Quality standards ▪ Budget **Activities** ▪ Create framework and structure of responsi-bilities and tasks ▪ Arrange regular meetings of working group and team leaders ▪ Develop communication flowchart ▪ Develop quality management plan and identify quality profile elements ▪ Implement a certifi-cation process to check	▪ Monitor budget, costs, and timeline for each country	▪ Study goals and objectives ▪ Documentation and formatting requirements ▪ All study implementation procedures ▪ Documentation of modifications to study protocol ▪ Summary of each organization's performance

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
	study design and quality standards		
Ethical Considerations in Surveys	**Inputs** ▪ Standards for ethical and scientific conduct ▪ Local and national human subject regulations and legislation ▪ Ethical guidelines in project management and human resource management ▪ Voluntary informed consent protocol and procedures ▪ Procedures for ethics training of project staff ▪ Comprehensive plan for protection of confidentiality **Activities** ▪ Review and apply ethical standards, best practices, and relevant regulations and legislation in designing study and collecting and disseminating survey data ▪ Develop and apply knowledge of local customs and norms relevant for designing culturally-sensitive survey protocols ▪ Pretest consent protocol and forms to ensure comprehension ▪ Translate consent protocols and forms according to best practices for translation ▪ Assess respondent burden (overall and by subgroup, if appropriate) ▪ Train project staff on ethics ▪ Have project staff sign pledge of confidentiality ▪ Complete ethics review submission and maintain documentation	▪ Report on staff completion of ethics training ▪ Review the implementation of informed consent procedures (percent of cases reviewed, percent of cases failing to follow procedures, actions taken, etc.) ▪ Report on interview falsification (percent of cases reviewed, percent of reviewed cases falsified, subsequent actions taken, etc.) ▪ Report on any actual or potential breaches of confidentiality, security, or other adverse event, including any resulting changes to study protocol ▪ Report on any failures of statistical disclosure control	▪ Description of voluntary consent and confidentiality procedures ▪ Copies of materials provided to respondents as part of informed consent process ▪ Summary of respondent burden assessment ▪ Description of ethics training for project staff ▪ Summary of ethics committee review ▪ Summary of review of recorded interviews regarding the implementation of informed consent procedures ▪ Summary of falsification findings ▪ Summary of any reported actual or potential breaches of confidentiality ▪ Description of disclosure analysis methods and summary of findings

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
(Ethical Considerations in Surveys)	of submission materials ▪ Review recorded interviews and monitoring to assure adherence to informed consent procedures ▪ Monitor implementation of confidentiality protocols and procedures ▪ Perform audits to determine adherence to confidentiality protocols and procedures ▪ Securely store signed pledges of confidentiality and consent forms ▪ Maintain records of all ethics review committee correspondence ▪ Conduct verification to detect possible interview falsification ▪ Conduct disclosure analysis		
Tenders, Bids, and Contracts	**Inputs** ▪ Type of contract offered ▪ Study specifications ▪ Minimum quality requirements and evaluation criteria for bids **Activities** ▪ Prepare tender based on study specifications ▪ Conduct competitive bidding process within each country ▪ Evaluate bids and select a survey organization in each country • Define progress approval points throughout the research process	▪ Report on evaluation scores of bidding organizations	▪ Summary of process of evaluating and selecting bidding organizations
Sample Design	**Inputs** ▪ Target and survey population descriptions ▪ Sampling frame(s), definitions, including definitions of strata and	▪ Estimate coverage error ▪ Report on percentage of duplicate and ineligible sampling units on the sampling frame(s) ▪ Produce tables/charts of	▪ Time dimension of design (e.g., one time cross sectional, fixed or rotating panel) ▪ Target and survey population definitions,

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
(Sample Design)	sampling units, and any updating of the frame that was needed ▪ Desired level of precision overall and for specific subgroups ▪ Sample size based on specified levels of precision ▪ Selection procedure(s) and estimates of probabilities of selection at each stage ▪ Field listing standard procedures and minimum requirements of field listers ▪ Unique, sample identification codes for each selected sampling unit ▪ Data dictionary of selected elements and sampling units with descriptive and distinct variable names and labels **Activities** ▪ Produce, update and/or clean sample frame(s), as needed ▪ Calculate sample size ▪ Implement selection procedure(s) ▪ Create a unique, sample identification code for each selected element or unit ▪ Arrange regular meetings of working group, project manager and sampling statistician ▪ Conduct responsive design plans to minimize survey costs and errors	paradata indicators that serve as proxies of survey costs and errors ▪ Alter the survey design during data collection to minimize costs and errors	including inclusion/ exclusion criteria ▪ Sampling frame(s) descriptions ▪ Examples of maps and protocol used in field listing ▪ Description of all stages of selection, including sample sizes, stratification, clustering, oversampling and number of replicates fielded at each stage ▪ Documentation of procedures to determine probabilities of selection and weights for each stage of selection ▪ Tables of the precision of the estimates of key survey statistics ▪ (If necessary), descriptions of substitution procedures
Questionnaire	**Inputs**		▪ Report on modifications

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
Design	▪ Research question ▪ Review of literature and any relevant studies to identify useful material ▪ Documentation templates ▪ Documentation of origin of any existing questions or materials to be considered for re-use **Activities** ▪ Create cross-cultural and cross-competence development team, providing briefing, training, and tools as relevant ▪ Determine design approach ▪ Create analysis plan relating constructs, indicators and question topics ▪ Implement design steps		made to questions at different stages
Adaptation	**Inputs** ▪ Source questionnaires and any materials which might be adapted ▪ Translated questionnaires and any materials which might be adapted ▪ Documentation templates as relevant ▪ Guidelines on adaptation goals and more common forms ▪ Briefing and training of team as necessary ▪ Delivery schedule and required outputs **Activities** ▪ Determine stage(s) at which adaptation is possible ▪ Create adaptation team with skills suited for whichever stage(s) are envisaged ▪ Make adaptation proposals with		

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
	documented justifications ▪ Conduct external review of adaptation proposals and their documentation ▪ Test adaptations for targeted population(s) and revise as relevant ▪ Adjudicate/sign-off on adaptation decisions and finalize documentation		
Translation	**Inputs** ▪ Source questionnaire and any material to be translated ▪ Guidelines and stipulations on procedures to be followed and on outputs required (e.g., need for documentation on decisions) ▪ Templates for translation development, as relevant ▪ Delivery schedule including any further refinements proposed that relate to translation (procedures such as language harmonization, adaptation, pretesting and any required adjudication steps **Activities** ▪ Create translation team, briefing, training and monitoring as relevant. ▪ Produce draft translations, checking translator output at an early stage of production ▪ Maintain documentation at each stage ▪ Review and adjudicate translations ▪ Pretest translations Repeat any translation refinement step as	▪	▪

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
	needed		
Instrument Technical Design	**Inputs** ▪ Instrument specification guidelines ▪ Comprehensive design evaluation plan, including goals, evaluation techniques, and timeline ▪ Quality assurance metrics (e.g., questionnaire and item timings, review of computer-assisted application audit trails, behavior/event codes) **Activities** ▪ Provide clear instrument specifications and/or data dictionary ▪ Provide culture or language-specific adaptations of design specifications ▪ Develop instrument evaluation procedures Perform and report on design assessments ▪ Review quality assurance metrics reports ▪ Make recommendations for improvement	▪ Collect and report on quality metrics or measures, such as: ▪ Questionnaire length and section and item timings ▪ Audit trails for computerized applications ▪ Behavior codes or event codes based on audio or video recordings of pretests or usability tests ▪ Qualitative analysis of cognitive and usability testing (see Pretesting) ▪ Heuristic evaluation or expert review	▪ Instrument specification guidelines ▪ Procedures for design evaluation ▪ Results of design evaluations ▪ Documentation and results of quality assurance and quality monitoring and control
Interviewer Recruitment and Training	**Inputs** ▪ Recruitment and training timeline ▪ Minimum standards for employment ▪ Study-specific requirements (e.g., gender, language, etc.) ▪ Assessment tests ▪ Minimum interviewer requirements checklist ▪ Criteria for dismissal or follow-up training ▪ Standard certification procedures **Activities** ▪ Train trainers before they train interviewers ▪ Complete checklist during candidate	▪ Report on training attendance ▪ Report on candidate training certification (including rates) ▪ Report on follow-up training certification (including rates)	▪ Employment criteria ▪ General and study-specific training documentation ▪ Certification procedures ▪ Certification rates for training and follow-up training

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
	screening ▪ Take attendance during training ▪ Certify candidates ▪ Dismiss or retrain candidates who fail certification ▪ Maintain written records of results of candidates' certification tests		
Pretesting	**Inputs** ▪ Pretesting plan, including pretest goals, evaluation techniques, timeline, and budget ▪ Standard procedures for staff training **Activities** ▪ Provide staff training and certification ▪ Review recordings of focus groups and cognitive interviews for staff errors ▪ Provide retraining as necessary ▪ Test for inter-coder reliability if appropriate	▪ Monitor costs and timeline ▪ Monitor staff error rates ▪ Test inter-coder reliability	▪ Pretest procedures documentation ▪ Pretest training documentation ▪ Pretest findings, change recommendations, and changes made ▪ Staff error rates
Data Collection	**Inputs** ▪ Target outcome rates (e.g., response, refusal, noncontact), and completion rates ▪ Target hours per interview ▪ Recontact or reinterview respondents ▪ Percentage of interviewer cases to be verified ▪ Verification questions ▪ Verification of case disposition codes and selected responses ▪ Interviewer performance checklist ▪ Criteria for interviewer dismissal or supplementary training **Activities** ▪ Establish a sample	▪ Overall, by key respondent groups and by interviewer, report on: ▪ Screening rates ▪ Eligibility rates ▪ Response rates ▪ Refusal rates ▪ Noncontact rates ▪ Completion rates ▪ Hours per interview ▪ Number of completed interviews ▪ Report on interviewer performance outcomes ▪ Develop a responsive design based on cost/error tradeoffs	▪ Documentation of mode(s) of data collection and the protocol for determining mode(s) to use ▪ Documentation of the sample management system ▪ Study materials ▪ Screening/respondent selection procedures ▪ Number of completed interviews, overall and by mode ▪ Documentation of proxy interview protocol ▪ Documentation of respondent incentives, and interviewer incentive protocol ▪ Documentation of techniques to maximize response (e.g., prenotification, recontact, and refusal conversion protocol)

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
	management system ▪ Review paper coversheets and/or questionnaires ▪ Dismiss or retrain interviewers with substandard performance ▪ Collect paradata needed for statistical adjustment		▪ Outcome rates, overall and by key respondent groups ▪ Dates of data collection ▪ Interviewer monitoring procedures and outcomes ▪ Verification form(s) and outcomes ▪ Any descriptions and outcomes of validation study (e.g., administrative record check against survey data)
Data Harmonization	**Inputs** ▪ Standard codebook specifications ▪ Standard procedures for collecting and producing national data files ▪ Comprehensive plan for harmonization of cross-cultural data files ▪ Procedures for testing harmonized files with knowledgeable users **Activities** ▪ Create cross-cultural monitoring team ▪ Periodically review analytic results to allow for changes in harmonization rules ▪ Review end-user test results ▪ Make recommendations for harmonization process improvement	▪ Report on analytic results ▪ Report on user tests	▪ Documentation of specification and procedures standards ▪ Documentation of conversion and harmonization decisions ▪ Results of user tests
Data Processing and Statistical Adjustment	**Inputs** ▪ Percent of manually entered questionnaires to be verified ▪ Criteria for data entry staff dismissal or supplementary training ▪ Items to be coded ▪ Coding protocol (manual or automatic) ▪ Percent of manually coded cases to be check coded ▪ Minimum acceptable	▪ Report on data entry accuracy rate ▪ Test inter-coder reliability Key process statistics for editing ▪ Edit failure rate ▪ Recontact rate ▪ Correction rate	Data processing ▪ Data coding and data entry training documentation ▪ Evaluation protocol for data coding and data entry staff and outcomes ▪ Items that were coded or re-coded ▪ Coding reliability ▪ Data entry verification protocol and outcomes ▪ Data editing protocol

Guidelines Chapter	Quality Planning and Assurance – Inputs and Activities	Quality Monitoring and Control – Measures and Reports	Elements of Quality Profile
	inter-coder reliability ▪ Data editing protocol ▪ Appropriate statistical software ▪ Appropriate statistical adjustments (e.g., imputation, weights) ▪ Appropriate standard error estimation ▪ Quality control procedures for calculation of statistical adjustments and variance estimation **Activities** ▪ Train data entry and data coding staff ▪ Verify data accuracy ▪ Develop coding scheme(s) ▪ Assess inter-coder reliability ▪ Check outliers ▪ Edit data		Statistical adjustment ▪ Rationale for assigning sample identification numbers ▪ Calculation of outcome rates (e.g., response, refusal, noncontact), weighted and unweighted ▪ Standard error estimates ▪ Percent item missing data Where applicable: ▪ Imputation method(s) ▪ Generation of weight(s) ▪ Trimming of weight(s) ▪ Scaling of weight(s) ▪ Adjustment(s) for differential nonresponse ▪ Poststratification adjustment(s)
Data Dissemination	**Inputs** ▪ Procedures for testing accessibility of archives with knowledgeable users ▪ Procedures for electronic preservation of files ▪ Procedures for testing files with major statistical packages **Activities** ▪ Create electronic versions of all files ▪ Provide data files in all major statistical software packages ▪ Designate resources to provide user support and training for secondary researchers ▪ Review results of user tests	▪ Data archive test reports	▪ Description and classification of target users and their needs ▪ Results of user satisfaction assessments ▪ Summary of conditions of access to data, accompanying documentation, and user feedback ▪ Distribution reports (dataset requests, Web hits, downloads, etc.)

Glossary

Accuracy	The degree of closeness an estimate has to the true value.
Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).
Adjustment Error	Survey error (variance and bias) due to post data collection statistical adjustment.
Audit trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.
Auxiliary data	Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in imputation techniques.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.

Bid	A complete proposal (submitted in competition with other bidders) to execute specified jobs within prescribed time and budget, and not exceeding a proposed amount.
Cause and effect diagram	A fishbone-structured diagram for a process, used as a brainstorming tool to help understand or improve the process. The main bone represents the process (e.g., interviewer training), and bones coming off of the main bone are pre-identified factors (e.g., training materials) that may affect the quality of the process. From there potential causes (lack of resources and time) and effects (poor quality materials) can be discussed, and solutions identified. Also known as a fishbone or Ishikawa diagram.
Certification	Objective assessment of performance. Based on pre-established criteria, the interviewer either meets the requirements and may proceed to conduct the study interview or does not meet the requirements and may either be permitted to try again or be dismissed from the study. Certification outcome should be documented and filed at the data collection agency.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Codebook	A document that provides question-level metadata that is matched to variables in a dataset. Metadata include the elements of a data dictionary, as well as basic study documentation, question text, universe statements (the characteristics of respondents who were asked the question), the number of respondents who answered the question, and response frequencies or statistics.
Coding	Translating nonnumeric data into numeric fields.
Coefficient of Variation (CV)	The ratio of the standard deviation of a survey estimate and its mean value. Its purpose is to cancel the unit of measurement and create a relative measure of variation that facilitates comparisons across different statistics.

Cognitive interview	A pretesting method designed to uncover problems in survey items by having respondents think out loud while answering a question or retrospectively.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Construct validity	The degree to which a survey question adequately measures an intended hypothetical construct. This may be assessed by checking the correlation between observations from that question with observations from other questions expected on theoretical grounds to be related.
Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Context effects	The effect of question context, such as the order or layout of questions, on survey responses.

Conversion process	Data processing procedures used to create harmonized variables from original input variables.
Cooperation rate	The proportion of all elements interviewed of all eligible units ever contacted.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Copyeditor	The person who reviews a text and marks up any changes required to correct style, punctuation, spelling, and grammar errors. In many instances, the copyeditor may also make the corrections needed.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
Coverage error	Survey error (variance and bias) that is introduced when there is not a one-to-one correspondence between frame and target population units. Some units in the target population are not included on the sampling frame (undercoverage), some units on the sampling frame are not members of the target population (out-of-scope), more than one unit on the sampling frame corresponds to the same target population unit (overcoverage), and one sampling frame unit corresponds to more than one target population unit.
Coversheet	Electronic or printed materials associated with each element that identify information about the element, e.g., the sample address, the unique identification number associated with an element, and the interviewer to whom an element is assigned. The coversheet often also contains an introduction to the study, instructions on how to screen sample members and randomly select the respondent, and space to record the date, time, outcome, and notes for every contact attempt.
Crosswalk	A description, usually presented in tabular format, of all the relationships between variables in individual data files and their counterparts in the harmonized file.

Data dictionary	A document linking the survey instrument (questionnaire) with the dataset, or more abstract question or variable-level metadata including question identifiers (variable names and labels); response category identifiers (value labels), and data types (e.g., F2.0, specifying that the response is a two-digit integer with zero decimal places.
Disclosure analysis and avoidance	The process of identifying and protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. For any proposed release of tabulations or microdata, the level of risk of disclosure should be evaluated.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Eligibility Rate	The number of eligible sample elements divided by the total number of elements on the sampling frame.
Ethics review committee or human subjects review board	A group or committee that is given the responsibility by an institution to review that institution's research projects involving human subjects. The primary purpose of the review is to assure the protection of the safety, rights and welfare of the human subjects.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."

Fixed panel design	A longitudinal study which attempts to collect survey data on the same sample elements at intervals over a period of time. After the initial sample selection, no additions to the sample are made.
Flow chart	A method used to identify the steps or events in a process. It uses basic shapes for starting and ending the process, taking an action, making a decision, and producing data and documentation. These are connected by arrows indicating the flow of the process. A flow chart can help identify points at which to perform quality assurance activities and produce indicators of quality that can be used in quality control.
Focus group	Small group discussions under the guidance of a moderator, often used in qualitative research that can also be used to test survey questionnaires and survey protocols.
Hours Per Interview (HPI)	A measure of study efficiency, calculated as the total number of interviewer hours spent during production (including travel, reluctance handling, listing, completing an interview, and other administrative tasks) divided by the total number of interviews.
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Item nonresponse, item missing data	The lack of information on individual data items for a sample element where other data items were successfully obtained.
Listing	A procedure used in area probability sample designs to create a complete list of all elements or cluster of elements within a specific set of geographic boundaries.
Longitudinal study	A study where elements are repeatedly measured over time.
Majority country	A country with low per capita income (the majority of countries).

Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Metadata	Information that describes data. The term encompasses a broad spectrum of information about the survey, from study title to sample design, details such as interviewer briefing notes, contextual data and/or information such as legal regulations, customs, and economic indicators. Note that the term 'data' is used here in a technical definition. Typically metadata are descriptive information and data are the numerical values described.
Microdata	Nonaggregated data that concern individual records for sampled units, such as households, respondents, organizations, administrators, schools, classrooms, students, etc. Microdata may come from auxiliary sources (e.g., census or geographical data) as well as surveys. They are contrasted with macrodata, such as variable means and frequencies, gained through the aggregation of microdata.
Mode	Method of data collection.
Noncontact	Sampling units that were potentially eligible but could not be reached.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Nonresponse bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.
Nonresponse error	Survey error (variance and bias) that is introduced when not all sample members participate in the survey (unit nonresponse) or not all survey items are answered (item nonreponse) by a sample element.

Nonresponse followup	A supplemental survey of sampled survey nonrespondents. Nonresponse followup surveys are designed to assess whether respondent data are biased due to differences between survey respondents and nonrespondents.
Outcome rate	A rate calculated based on the study's defined final disposition codes that reflect the outcome of specific contact attempts before the unit was finalized. Examples include response rates (the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.), cooperation rates (the proportion of all units interviewed of all eligible units ever contacted), refusal rates (the proportion of all units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units), and contact rates (the proportion of all units are reached by the survey).
Outlier	An atypical observation which does not appear to follow the distribution of the rest of a dataset.
Paradata	Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.
Pareto chart	A bar chart that reflects the types of most errors in a process, by error type in descending order; for example, the five or six most frequent types of help desk calls from interviewers using computer-assisted interviewing.
Performance measurement analysis	A technique used in quality control to determine whether quality assurance procedures have worked. For example, analysis of routine measures of interviewer or coder performance.

Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins.
Pledge of confidentiality	An agreement (typically in written or electronic form) to maintain the confidentiality of survey data that is signed by persons who have any form of access to confidential information.
Poststratification	A statistical adjustment that assures that sample estimates of totals or percentages (e.g., the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g., the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection.
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Probability sampling	A sampling method where each element on the sampling frame has a known, non-zero chance of selection.

Process analysis	The use of tools such as flowcharts to analyze processes, e.g., respondent tracking, computerized instrument programming and testing, coding, data entry, etc. The aim is and to identify indicators or measures of the quality of products. Process analysis also is used to identify improvements that can be made to processes.
Process improvement plan	A plan for improving a process, as a result of process analysis. A process improvement plan may result from development of a quality management plan, or as a result of quality assurance or quality control.
Process indicator	An indicator that refers to aspects of data collection (e.g., HPI, refusal rates, etc.).
Processing error	Survey error (variance and bias) that arise during the steps between collecting information from the respondent and having the value used in estimation. Processing errors include all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in the implementation of systems and estimates. In survey data, processing errors may include errors of transcription, errors of coding, errors of data entry, errors in the assignment of weights, errors in disclosure avoidance, and errors of arithmetic in tabulation.
Proxy interview	An interview with someone (e.g., parent, spouse) other than the person about whom information is being sought. There should be a set of rules specific to each survey that define who can serve as a proxy respondent.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.

Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Quality profile	A comprehensive report prepared by producers of survey data that provides information data users need to assess the quality of the data.
Recontact	To have someone other than the interviewer (often a supervisor) attempt to speak with the sample member after a screener or interview is conducted, in order to verify that it was completed according to the specified protocol.
Refusal rate	The proportion of all units of all potentially eligible sampling units in which a respondent sampling unit refuses to do an interview or breaks off interviews of all potentially eligible sampling units.
Reinterview	The process or action of interviewing the same respondent twice to assess reliability (simple response variance).
Reliability	The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.
Replicates	Systematic probability subsamples of the full sample.

Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Restricted-use data file	A file that includes information that can be related to specific individuals and is confidential and/or protected by law. Restricted-use data files are not required to include variables that have undergone coarsening disclosure risk edits. These files are available to researchers under controlled conditions.
Reviewer	Person who participates in the review of translations in order to produce a final version (see Appendix A of Translation).
Rotating panel design	A study where elements are repeatedly measured a set number of times, then replaced by new randomly chosen elements. Typically, the newly-chosen elements are also measured repeatedly for the appropriate number of times.
Sample design	Information on the target and final sample sizes, strata definitions and the sample selection methodology.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sample management system	A computerized and/or paper-based system used to assign and monitor sample units and record documentation for sample records (e.g., time and outcome of each contact attempt).
Sampling error	Survey error (variance and bias) due to observing a sample of the population rather than the entire population.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.

Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Source questionnaire	The questionnaire taken as the text for translation.
Statistical process control chart	A statistical chart that compares expected process performance (e.g., number of hours worked by interviewers in a week) against actual performance. For example, interviewers who perform outside upper and lower boundaries on this measure are flagged; if greater variation from expected performance for some interviewers in a certain location can be explained (e.g., a hurricane or a snow storm causing lower than expected hours worked), the process is in control; if not, corrective actions are taken.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Stratification	A sampling procedure that divides the sampling frame into mutually exclusive and exhaustive groups (or strata) and places each element on the frame into one of the groups. Independent selections are then made from each stratum, one by one, to ensure representation of each subgroup on the frame in the sample.

Substitution	A technique where each nonresponding sample element from the initial sample is replaced by another element of the target population, typically not an element selected in the initial sample. Substitution increases the nonresponse rate and most likely the nonresponse bias.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Task	An activity or group of related activities that is part of a survey process, likely defined within a structured plan, and attempted within a specified period of time.
Tender	A formal offer specifying jobs within prescribed time and budget.
Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Translator	The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other tasks such as reviewing and copyediting.
Unique Identification Number	A unique number that identifies an element (e.g., serial number). That number sticks to the element through the whole survey lifecycle and is published with the public dataset. It does not contain any information about the respondents or their addresses.

Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Universe statement	A description of the subgroup of respondents to which the survey item applies (e.g., “Female, ≥ 45, Now Working”).
Usability testing	Evaluation of a computer-assisted survey instrument to assess the effect of design on interviewer or respondent performance. Methods of evaluation include review by usability experts and observation of users working with the computer and survey instrument.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.
Working group	Experts working together to oversee the implementation of a particular aspect of the survey lifecycle (e.g., sampling, questionnaire design, training, quality control, etc.)

References

[1] Aitken, A., Hörngren, J., Jones, N., Lewis, D., & Zilhäo, M. J. (2003). *Handbook on improving quality by analysis of process variables.* Luxembourg: Eurostat. Retrieved March 27, 2010, from http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20ON%20IMPROVING%20QUALITY.pdf

[2] Anderson, R., Kasper, J., Frankel, M., & Associates (Eds.). (1979). *Total survey error: Applications to improve health surveys*. San Francisco: Jossey-Bass.

[3] Australian Bureau of Statistics. (2009). *Draft report on quality assurance frameworks*. Retrieved May 7, 2010, from http://unstats.un.org/unsd/dnss/qaf/qafreport.htm

[4] Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.

[5] Brackstone, G. (1999). Managing data quality in a statistical agency. *Survey methodology, 25*(2), 1-23.

[6] Cochran, W. G. (1977). *Sampling techniques*. New York, NY: John Wiley & Sons.

[7] Couper, M. P. (1998). *Measuring survey quality in a CASIC environment*. Paper presented at the Proceedings of American Statistical Association: Survey Research Methods Section. Retrieved March 27, 2010, from http://www.amstat.org/sections/srms/proceedings/papers/1998_006.pdf

[8] Couper, M. P., & Lyberg, L. E. (2005). *The use of paradata in survey research*. Paper presented at Proceedings of the 54th Section of International Statistical Institute.

[9] Defeo, J. A. & Juran , J. M. (2010). *Juran's quality handbook: The complete guide to performance excellence* (6th ed.). New York, NY: McGraw-Hill.

[10] Eurostat. (2003). *Methodological documents – Definition of quality in statistics* (Report of the Working Group Assessment of Quality in Statistics, item 4.2). Luxembourg: Eurostat. Retrieved March 27, 2010, from http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf

[11] Eurostat. (2003). *Methodological documents – Standard report* (Report of the Working Group Assessment of Quality in Statistics, item 4.2B). Luxembourg: Eurostat. Retrieved March 27, 2010, from http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/docume

nts/STANDARD_QUALITY_REPORT_0.pdf

[12] Eurostat. (2005). *Standard quality indicators* (Report to Metadata Working Group 2005,Doc. ESTAT/02/Quality/2005/9/Quality Indicators). Luxembourg: Eurostat. Retrieved May 7, 2010, from http://circa.europa.eu/Public/irc/dsis/metadata/library?l=/metadata_working_1/metadata_working_1/qualityindicatorspdf/_EN_1.0_&a=d

[13] Groves, R. M. & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*, 439–457.

[14] Groves, R. M. (1989). *Survey errors and survey costs*. Hoboken, NJ: John Wiley & Sons.

[15] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[16] International Monetary Fund (IMF). (2003). *Data quality assessment framework.* Retrieved May 7, 2010 from http://dsbb.imf.org/images/pdfs/dqrs_Genframework.pdf

[17] Lyberg, L. E., & Stukel, D. M. (2010). Quality assurance and quality control in cross-national comparative studies. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 227–250). Hoboken, NJ: John Wiley & Sons.

[18] Lyberg, L. E., & Biemer, P. P. (2008). Quality assurance and quality control in surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology*. New York/London: Lawrence Erlbaum Associates.

[19] Lyberg, L. E., Biemer, P. P., Collins, M., de Leeuw, E. D., Dippo, C., Schwarz, N. et al. (Eds.). (1997). *Survey measurement and process quality*. New York, NY: John Wiley & Sons.

[20] Lyberg, L. E., Bergdahl, M., Blanc, M., Booleman, M., Grünewald, W., Haworth, M. et al. (2001). *Summary report from the Leadership Group (LEG) on quality*. Retrieved March 1, 2010, from http://siqual.istat.it/SIQual/files/LEGsummary.pdf?cod=8412&tipo=2

[21] Lynn, P. (2003) Developing quality standards for cross-national survey research: five approaches. *International Journal of Social Research Methodology, 6*(4), 323-336. Retrieved October 8, 2009, from

http://www.soc.uoc.gr/socmedia/papageo/developing%20quality%20standards%20for%20cross-national%20surveys.pdf

[22] Lynn, P. (Ed.) (2006). *Quality profile: British Household Panel Survey (Version 2.0).* Colchester: Institute for Social and Economic Research, University of Essex. Retrieved March 27, 2010, from . http://www.iser.essex.ac.uk/files/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf

[23] Morganstein, D., & Marker, D. A. (1997). Continuous quality improvement in statistical agencies, in L. E. Lyberg, P. P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin. (Eds.), *Survey measurement and process quality*, New York: John Wiley & Sons.

[24] Mudryk, W., Burgess, M. J., & Xiao, P. (1996). Quality control of CATI operations in Statistics Canada. *Proceedings of the section on survey research methods* (pp.150-159). Alexandria, VA: American Statistical Association.

[25] Pierchala, C., & Surti, J. (1999), *Control charts as a tool in data quality improvement* (Technical Report No. DOT HS 809 005). Washington, DC: National Highway Traffic Safety Administration (NHTSA). Retrieved March 27, 2010, from http://www-nrd.nhtsa.dot.gov/Pubs/809005.PDF

[26] Project Management Institute. (2004). *A guide to the project management body of knowledge (PMBOK® guide)* (3rd ed.). Newtown Square, PA: Project Management Institute, Inc.

[27] Statistics Canada. (2002). *Statistics Canada quality assurance framework*. Ottawa: Statistics Canada. Retrieved May 7, 2010, from http://www.statcan.gc.ca/pub/12-586-x/12-586-x2002001-eng.pdf

[28] United States Bureau of the Census. (2006). *Definition of data quality*, (V1.3). Washington, DC: U.S. Bureau of the Census. Retrieved May 7, 2010, from http://www.census.gov/quality/P01-0_v1.3_Definition_of_Quality.pdf

[29] United States Bureau of the Census. (1998). *Survey of income and program participation: SIPP quality profile* (Working Paper No. 230) (3rd ed.).. Washington, DC: U.S. Bureau of the Census. Retrieved March 27, 2010, from http://www.census.gov/sipp/workpapr/wp230.pdf

III. Ethical Considerations in Surveys

Kirsten Alcser, Christopher Antoun, Ashley Bowers, Judi Clemens, and Christina Lien

Introduction

These guidelines focus on ethical concerns with regard to cross-cultural surveys as human subject research. The World Health Organization defines human subject research as the "…systematic collection or analysis of data…in which human beings (i) are exposed to manipulation, intervention, observation, or other interaction with investigators either directly or through alteration of their environment, or (ii) become individually identifiable through investigators' collection, preparation, or use of biological material or medical or other records" [25].

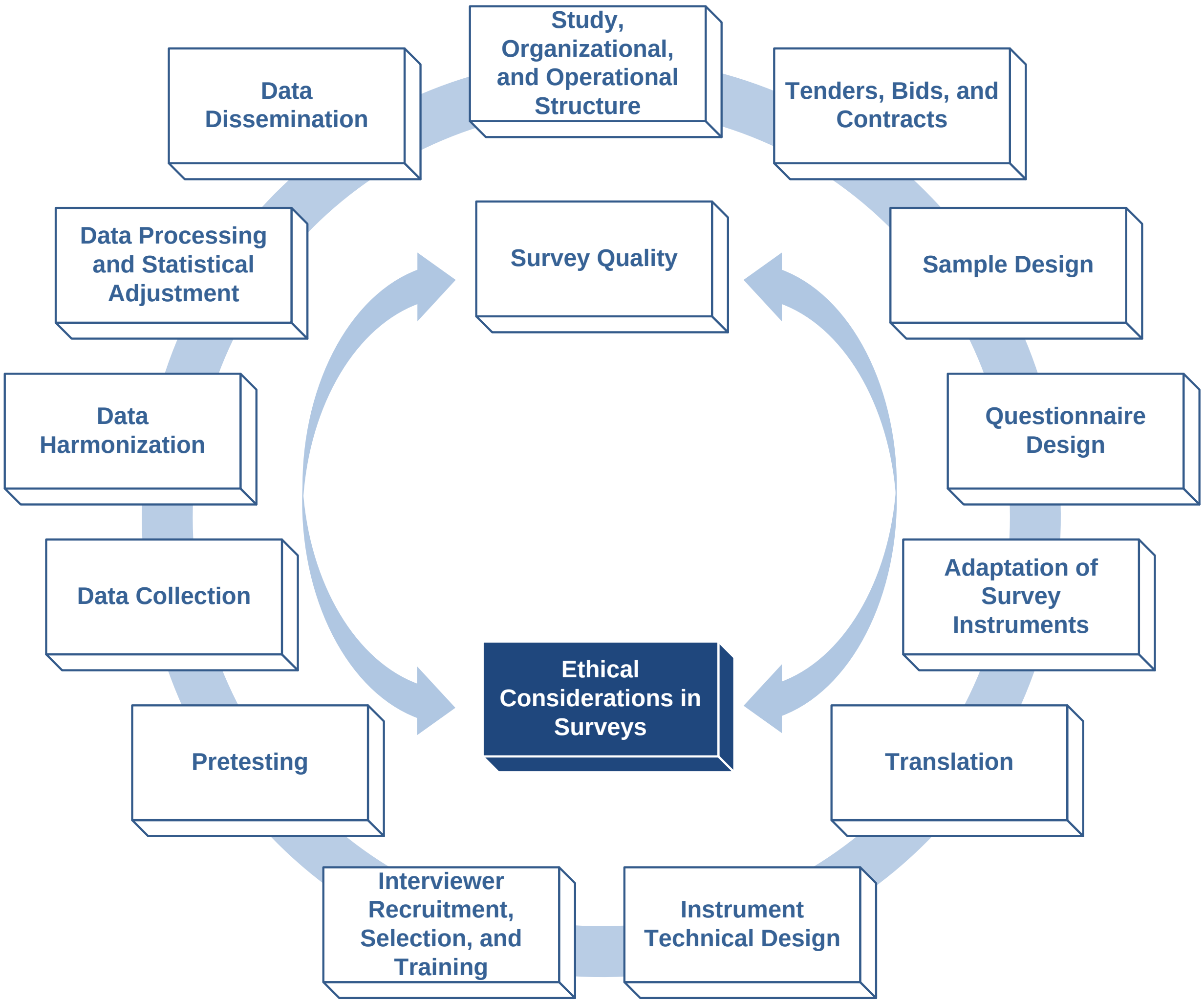
There is no lack of source material on ethical guidelines for human subject research (see [22], for a recent review). For example, the Declaration of Helsinki [26], originally adopted by the World Medical Association in 1964 and most recently revised in 2004, defines the ethical responsibilities of physicians to their patients and to the subjects of biomedical research. The principles in the Declaration of Helsinki have been extended to include social science human subject research. Professional organizations, such as the American Association for Public Opinion Research (AAPOR), the World Association for Public Opinion Research (WAPOR), the European Society for Market Research (ESOMAR), and the International Statistical Institute (ISI), have also developed ethical codes and guidelines for their members.

In addition to these self-regulatory measures, many countries have legislation in place that affects human subject research (e.g., data protection legislation and requirements for ethics review boards). Whether working in familiar surroundings or in new contexts, researchers must make sure they are informed about, and comply with, relevant legislation. When working in other countries or locations, researchers may need to comply not only with local requirements, pertaining to the place where they are collecting data, but also with their own country's requirements. A compilation of laws, regulations and guidelines from 96 countries has been prepared by the US Office for Human Research Protections and can be found on the Internet [24].

As might be expected, there is considerable overlap in the principles contained in the various ethics codes, professional association guidelines, and government regulations. This section attempts to consolidate their common elements, as well as to highlight concerns particular to cross-cultural studies, including cross-national variation in laws and regulations relevant to human subject research and cultural differences that affect the conduct of ethical research across cultures. It is important to recognize that researchers may confront tradeoffs between ethical principles. For example, maintaining sensitivity to cultural differences by having other family members present during the interview may conflict with ethical obligations to protect confidentiality and to minimize error in respondent reporting. For further information on the ethical principles presented here, please see the listing of ethics codes, declarations, guidelines, and other resources for researchers conducting cross-cultural human subject research that is provided in the Further Reading section.

Figure 1 shows ethical considerations within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To ensure that participating research teams follow widely accepted standards for ethical, professional, and scientific conduct from the design of the study through implementation, dissemination, and reporting.

1. **Protect the rights of free will, privacy, confidentiality, and well-being of research participants, and minimize the burden of study participation to the greatest extent possible.**

Rationale

The social researcher's responsibility to protect the human rights of study participants is universally prescribed in ethics codes and guidelines, such as the Declaration of Helsinki [26], and monitored by ethics review boards (in some countries). In addition, the collection of accurate data depends upon the cooperation of respondents: individuals are more likely to agree to participate in a study and to give complete, accurate information if they feel that they can trust the research organization. Finally, a positive experience with regard to the research interaction encourages participation in future research.

Procedural steps

- Avoid undue intrusion.
  - Use existing data whenever possible; do not collect new data unnecessarily.
  - Encourage participation in the research study only in ways that avoid personal harassment. This may include limiting the number of times that an interviewer visits a household to attempt to obtain sample member participation.
  - Be respectful and honest with survey respondents (e.g., be honest about the length of the interview, any benefits being offered, and the purpose of the study).
  - Adapt the study protocol as needed to protect the rights of participants from vulnerable populations, such as children, pregnant women, the elderly, prisoners, the mentally impaired, and members of economically and otherwise disadvantaged groups, by using special consent procedures (e.g., obtaining consent from a parent or family member) or other appropriate study modifications.
  - Keep respondent burden as low as possible [5].
    - Ensure that each question in the survey maps to a specific research goal.
    - Balance the need for information against the effort that is required to complete additional questions.
    - Ask questions in a way that is easy for respondents to answer (see [6], [7], and [9] for guidance).
    - If sensitive or otherwise demanding information is required, devise ways to help respondents provide it without undue burden. For example, part of the interview could be self-administered if there is concern that respondents might be uncomfortable providing responses to an interviewer.
  - Determine whether asking respondents to provide information on specific topics could bring harm or political repercussions to them and do not include questions on those topics.

- Consider carefully whether the requested information may be seen as private, threatening or embarrassing by the population interviewed, and implement techniques to minimize unease. In mental health studies, provisions are often made to provide suitable support for respondents or interviewers who experience emotional distress (for example, some form of emotional or psychological support service). In addition, interviewers in these studies complete specialized training on how to handle interviewing on sensitive topics. Also, recognize that cultures differ in what topics can be discussed and how they can be discussed.
- Consider the sensitivity of the requested information and assess whether a person other than the respondent would be able to provide the information in order to determine whether a proxy interview may be appropriate.
- If proxy interviews are used, create and adhere to a clearly defined set of rules defining who can serve as a proxy respondent**.**
- If the target respondent has indicated any unwillingness to provide information, do not gather the information from the proxy instead. Take care not to affect the relationship between the proxy and the target respondent.

- Obtain voluntary informed consent [10]. In implementing the consent process, provide the following information and adhere to the following principles.
  - Information to provide (in oral or written form, as appropriate):
    - A clear identification of the research firm affiliation.
    - A brief description of the survey.
    - A description of the role of the respondent in the study, including the expected duration of the respondent's participation.
    - An explanation of how the respondent was selected for the study.
    - A clear indication that participation is voluntary and that the information provided will be held confidential to the extent allowed by law (unless there are special circumstances in which respondents have waived confidentiality). It is important to note that in nearly all instances, respondents who are providing data to an interviewer cannot and should not be assured anonymity. In only rare instances generally involving self-administered surveys can respondents be promised that their data will be kept anonymous, that is, without any name or identifier ever associated with their response.
    - A clear description of any benefits and risks associated with participation.

  - Contact information for a study investigator or other research team member whom respondents can contact (provided or available on request).
  - If the study has been reviewed by an ethics review board, contact information for a review board member whom respondents can contact (provided or available on request).
- Principles to follow:
  - Do not use coercion. Whether a practice is defined as coercive or not may vary by culture, population, and study. Large monetary payments that are given to participants may be considered coercive in some studies.
  - Respect the rights of individuals to refuse to be interviewed, to refuse part of the interview, and to terminate an interview in progress. Whether or not follow-up with individuals who initially refuse the survey request is appropriate may vary by culture, population, and study.
  - Respect the right of individuals to refuse to answer any question in the interview.
  - Obtain and document consent. Whether consent is obtained in oral or written form depends on a number of factors, including government laws and regulations, risk of harm for respondents revealing sensitive information, the mode of data collection, the type of information requested, and cultural norms. In mail surveys, consent may be implied (that is, not explicitly obtained in oral or written form) if the respondent chooses to fill out the questionnaire and mail it back.
  - Obtain informed consent from a parent or responsible adult before interviewing children or young people.
  - Avoid making inaccurate or overly restrictive statements (e.g., the data will only be shared with the research team) if the data will be archived and shared with the research community [10].
- Consent information should be conveyed in a format that is easy for respondents to understand. Written formats that may be appropriate include a document with narrative text, a list of Frequently Asked Questions (FAQs), and a brochure format. Samples of these formats can be found in Appendix A and [3].

- Protect rights to privacy of study participants. This should include a careful review of government privacy laws and regulations, which could vary on the type of data and persons that are covered and the definition of an "identifiable" case [4].
  - Obtain the permission of respondents before using electronic equipment (e.g., taping, recording, photographing) and one-way viewing rooms.

- To the extent allowed by law or regulations, train staff to keep confidential both identifying material (e.g., respondent names, addresses, and phone numbers) and all information given by respondents.
- Require staff to sign a pledge of confidentiality or to provide assurance in some form that they will maintain confidentiality (see Appendix B for an example of a pledge of confidentiality). It is important to note that preserving confidentiality takes on even greater significance if local interviewers are working in areas where they may be acquainted with sample members prior to the interview request.
- Separate personally identifiable information (PII) from the respondent data. PII minimally includes name, address, phone number and identification number(s) (including an identification number assigned by a government agency such as a social security number in the United States or a driver's license number), but may include other information including biometric data.
- Keep secure and confidential any data source which links survey responses to identifiable respondents.
- Limit access to confidential data to project staff members who have pledged to maintain confidentiality and have been trained on appropriate use of study data.
- Use information gained through the research activity for study-related purposes only.
- Adhere to government laws and regulations on storage, retention, and dissemination of survey data.
- If appropriate, obtain a certificate of confidentiality or other legal document for protection from the requirement to release the identity of a respondent in a legal proceeding. Make clear to respondents the extent to which confidentiality is protected.
- If disclosing survey data to outside parties, require all subcontractors, consultants, and third parties to enter into an agreement to maintain respondent confidentiality. This agreement should include an explicit statement that the outside party cannot use contact information or any other information to recontact the respondent for any reason not directly related to the study (e.g., data cannot be used to approach respondents for a different study or for marketing purposes).
- Report any breach of confidentiality according to ethics review board policies and government regulations.

Lessons learned

- The manner in which research is conducted can shape a community's views positively or negatively on research topics, research institutions, and assumed or actual funders of the research.
  - Project Camelot was a U.S. Department of Defense research study designed to evaluate the Chilean masses' potential for revolutionary political action, and to determine the most effective means of counteracting that action. Participating Chilean social scientists were not told that the U.S. Department of Defense was funding the project and would ultimately receive the data. When Chilean researchers learned the facts, the study was cancelled. The image of the U.S. funders and U.S. research suffered greatly [2].
  - In 1974, psychologist Stanley Milgram conducted a study at Yale University [17]. Test subjects were told that they were part of an experiment on punishment and memory, and that they would act as "teachers." The "teacher" subjects were instructed by the experimenter to administer an electric shock to a "learner" if the latter failed to perform as required. Unbeknownst to the subject, the "learner" was one of the research team and deliberately gave many incorrect answers. The subject was ordered by the experimenter to give higher and higher intensity shocks to correct this poor performance. Although in fact no shocks were administered, the majority of subjects believed that they were actually administering electric shocks to the "learner." As a result, subjects experienced distress and tension during the experiment; several even had seizures. The unethical Milgram study was highly criticized after the event, and became a landmark in the effort to develop ethical guidelines for social science research [10]. However, while attacked from an ethics perspective, the Milgram study made a major contribution to research on obedience in social psychology. This study illustrates how it can be a challenge to balance the goals of science and ethical considerations.

- It is important to be truthful in describing the purpose of the study and the intended uses of study data.
  - In a study in India, dishonest interviewers were believed when they told respondents that survey participation would result in new schools, roads, and an electricity supply [2].
  - In the 1994 International Adult Literacy Survey (IALS), respondents in one country were told that they were participating in a pretest when in fact they were unknowingly providing data for the main study itself [12].

- Cross-cultural studies may involve the use of field research methods. Participant observation is a field research technique that involves becoming a trusted, yet temporary, participant in the community under study [23]. This temporary membership may lead to feelings of abandonment on the part of the participants. Possible solutions include maintaining honesty with the participants and community as well as providing the researched community with a final copy of the research results in the community's native language [20].

- Proper, ethical conduct may be simple and straightforward in one location but require multiple steps in another.
  - In Western cultures, simple parental consent may suffice when studying minors. In Mali, on the other hand, a medical research team that wanted to study children under 9 years of age who had been exposed to malaria first discussed the study with a group of village elders. Next, they convened focus group discussions with the heads of extended families. Then, they held similar discussions with mothers whose children might become part of the malaria study. Finally, they obtained the consent of the individual families involved [8].

- Regarding respondent burden and privacy, the duration and location of interviews has varied among established cross-cultural studies. Round 4 of the Afrobarometer Survey lasts approximately one hour and is usually administered in the respondent's home, although other locations are sometimes used [27]. Similarly, the Asian Barometer interview is completed in the respondent's home or workplace [28]. The basic face-to-face portion of the European Social Survey (ESS), Round 5, takes approximately 60 minutes and is conducted in the respondent's home [29]. The International Social Survey Programme (ISSP) questionnaire consists of 60 questions, not including demographics, and takes approximately 15 minutes to complete [30]. The length of the Living Standard Measurement Study Survey (LSMS) varies across participating countries, depending upon the number of modules administered [15]. The Survey of Health, Ageing and Retirement in Europe (SHARE) is completed in the respondent's home; it takes approximately 80 minutes to administer to a single-family household, and 120 minutes to administer to a multi-family household [31]. The average length of the interview for the World Mental Health Survey varies across participating countries, ranging from 49 minutes as a computer-assisted interview in Italy to 210 minutes as a paper-and-pencil interview in South Africa; most interviews are administered in the respondent's home, but in some countries, they are conducted in the respondent's place of employment, group quarters, cafes, libraries, or the office of the research organization [13].

2. Maintain sensitivity to cultural and social differences.

Rationale

Designing study protocols that are sensitive to cultural traditions and norms is vital to building trust and gaining cooperation. Being respectful of cultural norms and customs also leaves individual participants with a positive impression of the research community. Beyond the individual level, it may forestall negative political and social consequences. Finally, participation in social science and health studies may promote awareness of research issues in the community.

Procedural steps

- Do not exclude minority groups, native populations, or aboriginal peoples in the sample, unless it is appropriate to do so.

- Identify ethnic or religious power structures in the areas in which data collection will occur and approach study participants in accordance with the cultural traditions and norms of the ethnic or religious groups (e.g., through the head of the family or a local leader).

- Involve other individuals or groups in the consent decision-making process as appropriate (e.g., older family members or local leaders).

- Observe local customs in planning for and conducting the interview (e.g., giving advance notice before arriving, dressing in a culturally appropriate manner, removing one's shoes inside the house, partaking of refreshment, sending a thank-you note).

- Be flexible when implementing consent procedures (e.g., accepting oral consent in place of a written form, if literacy is an issue).

- Present study materials in a form that can be understood by the respondent (e.g., in the respondent's native language or orally rather than written if literacy is an issue). Avoid the use of technical language or jargon.

- Observe cultural norms when assigning interviewers to sample elements (e.g., matching female interviewers with female respondents, if matching is culturally appropriate).

- Attempt to conduct interviews in settings that afford as much privacy as possible while still respecting cultural norms (see Guideline 3 in Data Collection).

- Identify the level or degree of sensitivity for different question topics during preliminary fieldwork, observations, and pretesting, since sensitive topics often vary among cultures and societies [14].

- Consider cultural traditions and norms when deciding whether to offer respondent incentives and determining what type of incentives would be most appropriate (see Guideline 5 in Data Collection for more on incentives).

- Determine whether it is appropriate to follow up with persons who initially refuse the survey request and develop follow-up techniques in accordance with cultural traditions and norms.

Lessons learned

- As with other aspects of research, we cannot assume that "one size fits all" when implementing a study protocol with regard to ethics.
    - There may be different levels of requirements for privacy in different cultures. In a study involving 11-year-old boys in India, in-home interviews tended to include relatives and neighbors. At times the interviewers had to use considerable tact to discourage members of the audience from interjecting their own answers to the questions being asked [2].
    - In some cultures, it may be necessary to gain approval from authority figures within a community (gatekeepers). In a fertility study in Guatemala, interviewers were effectively barred from a rural municipality by the single act of a local priest. The priest warned his parishioners against the "red urbanites who would prevent women from having children," as he described the researchers [1].
    - Respondents in some cultures may be reluctant to provide written consent. Researchers in Mali found that documenting the consent process with a signed paper was a challenge. At first, villagers were opposed to signing any document, because they strongly believed that their word should be sufficient. In addition, participants found the legal language difficult to understand. It took very careful explanation and patience to overcome this resistance [8].

3. **Observe professional standards for managing and conducting scientifically-rigorous research at all stages of the study.**

Rationale

Researchers have a responsibility not only to protect participants but also to adhere to ethical management practices and to conduct research that meets the scientific standards of their field. The reader is referred here to other chapters which provide useful guidance on meeting scientific standards for the design, implementation, analysis, and documentation of cross-cultural surveys.

Procedural steps

- Clearly and objectively lay out the study's major research questions.

- Ensure that a survey is the most appropriate method to use to answer the research questions.

- Adhere to ethical business practices in bidding, contracting, and project management. These include the following:
  - Honestly describing the organization's expertise in a bid.
  - Disclosing if a survey project is being carried out on behalf of multiple clients or is using subcontractors.
  - Meeting contractual obligations.
  - Ensuring agreement by both parties on any changes to contractual obligations.
  - Maintaining good relations between the coordinating center and research organizations involved in the study.

  For additional detail, see Tenders, Bids, and Contracts.

- Disclose sources of financial support or relevant relationships that have the appearance of or potential to constitute a conflict of interest.

- Fulfill ethical responsibilities to employees (e.g., fair hiring practices, an objective performance evaluation process, and a commitment to employee safety). See Guideline 3 of the Data Collection chapter for guidance on the survey organization's responsibility to protect the well-being and safety of its interviewing staff.

- Train staff on the importance of ethics and scientific rigor in research involving human subjects (see other Guidelines in this chapter).
  - Ensure that interviewers are aware of their ethical responsibilities, including their obligation to report evidence of child abuse and other observations.
  - Instruct interviewers on the limits of their ethical responsibilities (e.g., when they should provide information about local health resources or contact a clinical psychologist or social worker

assigned to the project, rather than attempting to provide medical assistance or mental health support services themselves).

- Equip staff involved in design, data collection, and analysis with appropriate skills to perform scientifically rigorous research.

- Follow best practices in survey design, data collection, and post-survey processing as described in the following chapters:
  - Study, Organizational, and Operational Structure.
  - Survey Quality.
  - Tenders, Bids, and Contracts.
  - Sample Design.
  - Questionnaire Design.
  - Adaptation of Survey Instruments.
  - Translation.
  - Instrument Technical Design.
  - Interviewer Recruitment, Selection, and Training.
  - Pretesting.
  - Data Collection.
  - Data Harmonization.
  - Data Processing and Statistical Adjustment.
  - Data Dissemination.

- Employ appropriate tools and methods of analysis.

- Make interpretations of research results that are consistent with the data.

- Be clear and honest about how much confidence can be placed in the conclusions drawn from the data.

- Report research findings, even if they are not in line with the researcher's hypothesis.

- Monitor possible ethics violations, such as interviewer falsification or plagiarism, during the design, data collection, and analysis phases.

- Consider both cost and error implications of decisions that are made in the design, implementation, and analysis phases of the research study.

- When possible, conduct methodological studies to inform understanding of the cost and quality implications of survey design decisions for the benefit of future studies and the scientific research community. Most of the methodological research on ethics and other survey design considerations has been conducted in Western cultures. Additional research is needed in non-Western societies.

4. **Report research findings and methods and provide appropriate access to study data.**

Rationale

Professional social science organizations generally agree that their members should report findings to benefit the widest possible community. From this, it follows that data collection agencies should provide full information to allow readers and data users to assess both methodology and results. Dissemination of results and research reports also increases public confidence and alerts potential users to limits of accuracy and reliability, avoiding misinterpretation of findings. In addition, sharing documentation on study methods can assist other researchers in making informed choices about research design and implementation in future studies. While providing access to study data and methods is advantageous for the reasons outlined here, researchers must also assess the risk of a breach of confidentiality and address this concern when preparing data for dissemination.

Procedural steps

- Report findings as completely, widely and objectively as possible, while also protecting participants' confidentiality. While the full reporting of results is an important ethical obligation, it is also important to consider the negative impact that reporting unfavorable findings about a specific ethnic, religious, or other social group may have on members of that group.

- Make available as much of the study's methods, results, and raw data as possible, within the bounds of protecting participants' confidentiality, in order to permit others to evaluate the study and to replicate the findings.

- Evaluate the risk of a breach of confidentiality and implement appropriate techniques to protect the confidentiality of the data, including de-identification of publicly available datasets to the greatest extent possible (see Data Dissemination for a detailed discussion).

- Provide a summary report of the study methodology and findings. See Appendix C for a checklist of items to include in the summary report.

- Provide a copy of the findings to all researchers and organizations that were involved in the study.

- Provide a copy of the de-identified dataset(s) and documentation to international data repositories such as the Inter-university Consortium for Political and Social Research (ICPSR) [11], Council for European Social Science Data Archives (CESSDA), UK Data Archive (UKDA), or the South African Data Archive (SADA).

- Provide safe, sustainable storage of the datasets and documentation.

- Adhere to government laws and agreements that address disclosure of survey data both within and across borders.

- If an error is discovered after publication of the results, make an effort to correct the error using an erratum document that describes the error and its likely effect on study results, and provide an additional variable or other means along with appropriate documentation for analysts to identify the corrected value(s).

- Make an effort to respond to specific written requests for additional items pertaining to the publicly released findings [19].

Lessons learned

- There are useful examples of efforts to fully document study methods and provide survey data from cross-cultural studies to a wide community of users. In part or whole, their approach and templates can serve as models for other studies.
  - The European Social Survey website provides comprehensive information on study methodology and access to data for any registered user. Registration is free and easy to complete.
  - The World Mental Health Survey Initiative used a standardized web-based survey instrument to collect information on study methodology from participating countries.

- It is important to be aware that some national standards require that raw and de-identified datasets be stored for a minimum time period (e.g., 10 years is the German National Science Foundation standard for empirical data).

5. **Institute and follow appropriate quality control procedures.**

Rationale

Development and implementation of quality control procedures is necessary to ensure that the procedures that have been developed to meet standards for ethical research are being carried out appropriately. If a failure to meet these standards is detected, protocols should be in place to remedy the failure. In addition, monitoring of procedures related to the ethical conduct of the study should inform efforts to improve quality and cost-effectiveness.

Procedural steps

- Pretest consent protocol and forms to ensure comprehension.

- Translate and adapt consent protocols and forms according to best practices (see Translation and Adaptation of Survey Instruments).

- Review recorded interviews and monitor live interviews to assure adherence to informed consent procedures.
- Monitor implementation of confidentiality protocols and procedures, including, but not limited to performing audits to determine adherence to these protocols and procedures.

- Securely store signed pledges of confidentiality and consent forms.

- Recontact a sample of cases for each interviewer to verify that screening and interview procedures were appropriately followed (see Guideline 7 of Data Collection).

- Use analyses of paradata (e.g., identification of question-level timings that are unusually short or long and identification of unusual variable distributions for one or more interviewers compared to the overall distribution [18] [21]).

- Conduct disclosure analysis (see Data Dissemination).

- Investigate any deviation from ethical protocols and take appropriate action to address the situation.

Lessons learned

- Sometimes a small group of interviewers can have a large impact on the quality of survey estimates. In a mental health survey of six European countries, the prevalence rates of mental health disorders were unusually low among German respondents. Experienced

German interviewers were suspected of skipping screening questions that lead to a more extensive set of follow-up items in order to complete interviews more quickly. Even though only a small group of interviewers had prior interviewing experience, they conducted a sizeable percent of the total number of interviews and the responses that they solicited were very different. In general, positive responses screened respondents into more extensive sections on mental health disorders. Only 14.5% of screening questions administered by the interviewers with prior interviewing experience were positive, while 44.7% of screening questions administered by interviewers without prior experience were positive [16].

6. **Document materials and procedures related to the ethical conduct of the study and ethics committee reviews.**

Rationale

In research that involves human subjects, it is critical to maintain documentation of materials that were used to inform potential participants about study participation and subsequently record consent, in case there is ever a question of ethics violations or a request for additional information from an ethics review board. In addition, documentation of all survey procedures including those related to the ethical conduct of the study is a key element of high quality scientific research.

Procedural steps

- Maintain a copy of the following documents:
  - Scripts, letters, fact sheets, and any other materials provided to respondents to give them information they need to make an informed decision about participation.
  - Consent form templates and protocols.
  - Translated or adapted consent form templates and protocols.
  - Individual consent information for each respondent, stored in a safe environment separate from survey data.
  - Confidentiality procedures and protocols.
  - Pledge(s) of confidentiality completed by staff.
  - Records of completion of any specialized staff training on ethics.
  - Ethics review board original submission and requests for modification to study protocol (see Appendix D for a checklist of materials to include an ethics review board submission).
  - Ethics review board correspondence (e.g., letters of approval).
  - Any correspondence between study staff or ethics review board members/staff and respondents regarding an ethical issue or concern.

    - Reports of quality control activities (e.g., documentation of verification activities).

- Provide a copy of the following documents to any central coordinating organization:
    - Translated or adapted consent form templates and protocols.
    - Ethics review board original submission and requests for modification to study protocol.
    - Ethics review board correspondence (e.g., letters of approval).
    - Reports of quality control activities (e.g., documentation of verification activities).

Appendix A

Study brochure

The following is a sample study brochure that can be mailed or handed to respondents to provide general information about the study purpose and protocol and to address frequently asked questions.

What is this project about?

The Chicago Healthy Neighborhoods Study (CHNS) is a research study funded by the US National Office for Health to determine the impact of the quality of life in Chicago neighborhoods on the health of adults living there.

The information gathered from this study will help us better understand why there are social, economic, and racial/ethnic differences in the health of Chicagoans and how these differences affect Chicagoans' lives. With data from this study, effective approaches can be developed to improve the health and lives of all Chicagoans.

Who is asked to participate?

The CHNS is one of the largest surveys, done in a major American city, studying the relationship of the quality of people's lives and the neighborhood in which they live to their health. About 4,500 adults will participate in this important study.

Households are randomly selected using a scientific sampling procedure. Once a household is selected, an interviewer visits the house and makes a listing of all residents. One adult is randomly selected from all eligible residents. Only the selected individual may participate. Each person who is asked to participate has been carefully selected to represent fellow Chicagoans like them.

Is participation voluntary?

Yes. Participation in this project is voluntary. Project participants may choose not to answer any or all of the questions. However, each participant has been carefully selected and thus cooperation from each potential participant is critical to the success of this research.

How will the interviews be conducted?

Interviews will be conducted in the participant's home or at another location by a professional University of West Chicago Survey Research Center interviewer. The interviewer will ask questions and record answers using a laptop computer. Participants will be provided with $20 as a token of appreciation for their participation in this project.

What kinds of questions will I be asked?

The interview includes a wide range of questions about work and family life, health, and social and physical characteristics of neighborhoods in which study participants live. There are no right or wrong answers. Most participants find the interview to be an enjoyable experience.

How will the data be used?

The data collected will help researchers and government policy makers better understand social, economic, and racial/ethnic differences in the health of adults living in Chicago, so that effective approaches can be developed to improve the health and lives of all Chicagoans. Data from this study will only be reported in summary form. Participants' individual identities and answers to questions will remain strictly confidential.

Who is funding the project?

Funding for CHNS comes from the US National Office for Health (NOH).

The University of West Chicago's Survey Research Center will conduct the interviews for this study. A University of West Chicago interviewer will greet you at your home. For security reasons, you may want to ask the interviewer to reveal his/her identification badge. UWC employees will gladly comply with your request.

We thank you for your interest in this project!

If you have any questions, please contact the project team toll-free at:

1-800-733-7373

University of West Chicago Survey Research Center

Chicago Healthy Neighborhoods Study

Project Leader

- Christopher Antoun, Ph.D., Survey Research Center (SRC) & Department of Urban Health, University of West Chicago

Senior Investigators

- Benjamin Duffey, Ph.D., Department of Urban Health, University of West Chicago
- Hyun Jung Lee, Ph.D., Department of Urban Health & SRC, University of West Chicago
- Emily Blasczyk, Ph.D., Department of Psychology & SRC, University of West Chicago
- Mason Flounder, Ph.D., Department of Sociology, Northwestern University
- Yuchieh Lin, M.D., Mental Health Research Institute & Department of Psychiatry, University of West Chicago
- William Jones, M.D., Department of Psychiatry, University of West Chicago

Consultants

- Robert Kessenheimer, M.D., Department of Psychiatry, Loyola University (Chicago)
- Sara Neighbors, Ph.D., Department of Psychology, University of Pennsylvania

CHNS

CHICAGO HEALTHY NEIGHBORHOODS STUDY

Appendix B

Pledge to safeguard respondent privacy

This pledge to maintain respondent privacy is used by the Institute for Social Research at the University of Michigan. The form is signed by all staff members, and fulfillment of the pledge is a requirement of employment.

I have read the Institute for Social Research Policy on Safeguarding Respondent Privacy, and pledge that I will strictly comply with that Policy. Specifically:

- I will not reveal the name, address, telephone number, or other identifying information of any respondent (or family member of a respondent or other informant) to any person other than an employee directly connected to the study in which the respondent is participating.
- I will not reveal the contents or substance of the responses of any identifiable respondent or informant to any person other than an employee directly connected to the study in which the respondent is participating, except as authorized by the project director or authorized designate.
- I will not contact any respondent (or family member, employer, other person connected to a respondent or informant) except as authorized by the project director or authorized designate.
- I will not release a dataset (including for unrestricted public use or for other unrestricted uses) except in accordance with authorization, policies and procedures established by ISR and the Center with which I am affiliated.
- I will take all necessary precautions to avoid unintended disclosure of confidential information, including securing of paper and electronic records, computers, user IDs and passwords.

I agree that compliance with this Pledge and the underlying Policy is: 1) a condition of my employment (if I am an employee of ISR), and 2) a condition of continuing collaboration and association with ISR (if I am an affiliate of ISR). I understand that violation of this Policy and Pledge may result in disciplinary action, up to and including termination of employment or severance of any relationship with ISR and the applicable research project.

If I supervise affiliates who have access to ISR respondent data (other than unrestricted public release datasets), I will ensure that those affiliates adhere to the same standards of protection of ISR respondent privacy, anonymity, and confidentiality, as required by this Pledge and the associated Policy.

Signature: ______________________________ Date: ____________________
Typed or printed name: ____

Appendix C

Checklist of items to include in summary report of study methodology and findings

▪ The purpose of the study	☐
▪ Who sponsored the survey and who conducted it	☐
▪ A copy of ethics review board approval (if appropriate)	☐
▪ A copy of the informed consent form or script	☐
▪ A definition of the population under study and a description of the sampling frame	☐
▪ A description of the sampling and survey designs	☐
▪ Sample sizes and, where appropriate, eligibility criteria, screening procedures, and response rates. A summary of the disposition of sample elements should be included, in order for the user to calculate a response rate should one not be included in the report or a different one desired.	☐
▪ Method, location, and dates of data collection	☐
▪ A copy of questionnaire, interviewer instructions, and any visual aids used in the interview	☐
▪ A detailed description of results that are based on anything less than the total sample, including the size of the sample and inclusion/exclusion criteria	☐
▪ A full description of the weighting (if appropriate) and estimation procedures used for all results that are reported	☐
▪ The major findings	☐
▪ A description of the precision of the findings, including estimates of sampling error	☐

Primary Source: American Association for Public Opinion Research. (2005). *Standards for minimal disclosure.* Retrieved April 5, 2010, from http://www.aapor.org/Disclosure_Standards.htm

Appendix D

Checklist of materials to be provided to an ethics review board

General Study Information, including: • Financial sponsorship • Key personnel • Performance sites • Study dates • Study abstract/summary • Research design (including specific aims, background/prior research, methodology, analysis plan, etc.) • Benefits to subjects from participation • Risks to subjects • Recruitment methods and description of subject population • Informed consent procedures • Data confidentiality provisions • Conflicts of interest	☐
Discussion of Special Considerations, for example: • Procedures used to obtain consent to interview minors or other populations that require special consent (e.g., if interviewing minors, describe procedures for obtaining parental consent and include child assent and parental consent forms/oral protocols). • Compensation and costs involved in participation for study subjects • Procedures for handling biological samples, such as blood or saliva • Proposal to conduct genetic typing/analysis from biological samples • Considerations in conducting epidemiological or public health research • Use of deception • Use of internet/email for research	☐

• Consent procedures for audio or video recording of interviews • International research considerations • Protocols for viewing of images or listening to recorded material • Secondary data analysis	
Forms, including: • Copy of the grant/contract application • Consent protocols/scripts/forms • Copy of the questionnaire *Other forms (as appropriate):* • Cognitive interview protocol • Focus group moderator guide • Recruitment flyers or emails • Study brochure/fact sheet • Letter(s) to be sent to respondents • Data use agreement (for use of secondary data from third party sources) • Documentation of review from other ethics review boards • Documentation of training in research ethics for study staff	☐

Primary Source: University of Michigan Health Sciences and Behavioral Sciences Institutional Review Boards. Retrieved April 12, 2010 from http://www.irb.umich.edu/

Glossary

Accuracy	The degree of closeness an estimate has to the true value.
Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Anonymity	Recording or storing information without name or identifier, so the respondent cannot be identified in any way by anyone. No one can link an individual person to the responses of that person, including the investigator or the interviewer. Face-to-face interviews are never anonymous since the interviewer knows the address (and likely, the name) of the respondent.
Audit trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.
Auxiliary data	Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in imputation techniques.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Bid	A complete proposal (submitted in competition with other bidders) to execute specified jobs within prescribed time and budget, and not exceeding a proposed amount.

Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Cognitive interview	A pretesting method designed to uncover problems in survey items by having respondents think out loud while answering a question or retrospectively.
Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
De-identification	Separating personally identifiable information (PII) from the survey data to prevent a breach of confidentiality.

Disclosure analysis and avoidance	The process of identifying and protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. For any proposed release of tabulations or microdata, the level of risk of disclosure should be evaluated.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Ethics review committee or human subjects review board	A group or committee that is given the responsibility by an institution to review that institution's research projects involving human subjects. The primary purpose of the review is to assure the protection of the safety, rights, and welfare of the human subjects.
Fact sheet	A sheet, pamphlet, or brochure that provides important information about the study to assist respondents in making an informed decision about participation. Elements of a fact sheet may include the following: the purpose of the study, sponsorship, uses of the data, role of the respondent, sample selection procedures, benefits and risks of participation, and confidentiality.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Focus group	Small group discussions under the guidance of a moderator, often used in qualitative research that can also be used to test survey questionnaires and survey protocols.
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.

Informant	The person who supplies a list of the eligible elements within the selected unit. For example, many in-person surveys select a sample of housing units at the penultimate stage of selection. Interviewers then contact the housing unit with the aim of convincing the member of the housing unit who responded to the contact attempt to provide a list of housing unit members who are eligible for the study. The housing unit member who provides a list of all eligible housing unit members is called the informant. Informants can also be selected respondents as well, if they are eligible for the study and are chosen as the respondent during the within household stage of selection.
Interviewer falsification	Intentionally departing from the designed interviewer guidelines that could result in the contamination of the data. Falsification includes: 1) Fabricating all or part of an interview—the recording of data that are not provided by a designated survey respondent, and reporting them as answers of that respondent; 2) Deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a respondent refusal as ineligible for the sample; reporting a fictitious contact attempt); 3) Deliberately miscoding the answer to a question in order to avoid follow-up questions; 4) Deliberately interviewing a nonsampled person in order to reduce effort required to complete an interview; or intentionally misrepresenting the data collection process to the survey management.
Item nonresponse, item missing data	The lack of information on individual data items for a sample element where other data items were successfully obtained.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Microdata	Nonaggregated data that concern individual records for sampled units, such as households, respondents, organizations, administrators, schools, classrooms, students, etc. Microdata may come from auxiliary sources (e.g., census or geographical data) as well as surveys. They are contrasted with macrodata, such as variable means and frequencies, gained through the aggregation of microdata.
Mode	Method of data collection.

Noncontact	Sampling units that were potentially eligible but could not be reached.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Paradata	Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.
Personally Identifiable Information (PII)	Information that can be used to identify a respondent that minimally includes name, address, telephone number and identification number (such as social security number or driver's license number), but may include other information including biometric data.
Pledge of confidentiality	An agreement (typically in written or electronic form) to maintain the confidentiality of survey data that is signed by persons who have any form of access to confidential information.
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.

Proxy interview	An interview with someone (e.g., parent, spouse) other than the person about whom information is being sought. There should be a set of rules specific to each survey that define who can serve as a proxy respondent.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Recontact	To have someone other than the interviewer (often a supervisor) attempt to speak with the sample member after a screener or interview is conducted, in order to verify that it was completed according to the specified protocol or to edit potentially erroneous responses.
Reliability	The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.

Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sampling error	Survey error (variance and bias) due to observing a sample of the population rather than the entire population.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.

Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

References

[1] Amaro, N., & Gehlert Mata, C. (1968). Encuesta sobre el condicionamiento socio-cultural de la fecundidad en áreas marginales urbanas-metropolitanas, ladino-rurales e indígenas-tradicionales: Primera parte, marco teórico y metodológico [Survey on socio-cultural conditioning of fertility in urban marginal areas-metropolitan, rural and indigenous-ladino-traditional: Part One, theoretical and methodological framework] (2nd ed.). Guatemala: Instituto Centroamericano de Población y Familia.

[2] Armer, M., & Grimshaw, A. D. (1973). *Comparative social research*: *Methodological problems and strategies*. New York, NY: John Wiley & Sons.

[3] American Association for Public Opinion Research (AAPOR). (2010). *Examples of consent forms*. Retrieved May 7, 2010, from http://www.aapor.org/Consent_Form_Examples/1467.htm

[4] Benson, G. (2007). *The impact of privacy laws on survey data quality: A European case study.* Paper presented at the Council for Marketing and Opinion Research, Atlanta, GA.

[5] Bradburn, N. M. (1978). Respondent burden. *Proceedings of the section on survey research methods* (pp. 35-40). Alexandria, VA: American Statistical Association.

[6] Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage Publications.

[7] Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York, NY: John Wiley & Sons.

[8] Doumbo, O. K. (2005). It takes a village: Medical research and ethics in Mali. *Science, 307*(5710), 679–681. Retrieved May 13, 2008, from http://www.sciencemag.org/cgi/content/full/307/5710/679

[9] Fowler, F. J. Jr. (1995). *Improving survey questions: Design and evaluation* (*Vol. 38*, Applied social research methods series). Thousand Oaks, CA: Sage Publications.

[10] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Principles and practices related to ethical research. In R. M. Groves , F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer, & R. Tourangeau (Eds.), *Survey methodology* (2nd ed.) (pp. 371-400). Hoboken, NJ: John Wiley & Sons.

[11] Inter-university Consortium for Political and Social Research (ICPSR). (2009). *Guide to social science data preparation and archiving: Best practice throughout the data life cycle* (4th ed.). Ann Arbor, MI: ICPSR. Retrieved May 5, 2010, from http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

[12] Kalton, G., Lyberg, L. E., & Rempp, J. M. (1998). Review of Methodology. In S. Murray, I. Kirsch, and L. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first international adult literacy survey* (Report NCES 98-053). Washington, DC: U.S. Department of Education.

[13] Kessler, R. C., Ustun, T. B., & World Health Organization. (2008). *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press. Published in collaboration with the World Health Organization.

[14] Lee, R. M. (1993). Doing research on sensitive topics. London, UK: Sage Publications.

[15] Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

[16] Matschinger, H., Bernert, S., & Angermeyer, M. C. (2005). An analysis of interviewer effects on screening questions in a computer-assisted personal mental health interview. *Journal of Official Statistics, 21*(4), 657-674.

[17] Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18*(1), 57-76. Retrieved April 13, 2010, from http://hum.sagepub.com/cgi/reprint/18/1/57

[18] Murphy, J., Baxter, R. K., Eyerman, J., Cunningham, D., & Kennet, J. (2004). A System for Detecting Interviewer Falsification. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, Arizona. Retrieved April 12, 2010, from http://www.allacademic.com/meta/p116011_index.html

[19] National Council on Public Polls. (2006). *Principles of disclosure.* Retrieved January 29, 2010, from http://www.ncpp.org/files/NCPP%20POD%202006.pdf

[20] Punch, M. (1994). Politics and ethics in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research*

(pp. 83-95). Thousand Oaks, CA: Sage Publications.

[21] Schäfer, C., Schräpler, J. P., Müller, K. R., & Wagner, G. G. (2004). Automatic identification of faked and fraudulent interviews in surveys by two different methods. Paper presented at the European Conference on Quality and Methodology in Official Statistics, Mainz, Germany. Retrieved April 13, 2010, from http://www.diw.de/documents/publikationen/73/diw_01.c.42515.de/dp441.pdf

[22] Singer, E. (2008). Ethical issues in surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 78-96). New York/London: Lawrence Erlbaum Associates.

[23] Singleton Jr., R. A., & Straits, B. C. (2005). Approaches to social research (4th edition). New York, NY: Oxford University Press.

[24] United States Department of Health and Human Services Office for Human Research Protections. (2010). *US Office for Human Research Protections compilation of laws, regulations, and guidelines*. Retrieved May 4, 2010, from http://www.hhs.gov/ohrp/international/index.html#NatlPol

[25] World Health Organization (WHO). (2009). *Research leadership development training grants guidance for developing a research proposal*. Geneva: WHO. Retrieved October 7, 2009, from http://apps.who.int/tdr/svc/grants/calls/ltg/developing-research-proposal

[26] World Medical Association (WMA). (1964). *Declaration of Helsinki: Ethical principles for medical research involving human subjects* (adopted amended 29th WMA General Assembly, Tokyo, Japan, October 18th WMA General Assembly, Helsinki, Finland, June 1964; 1975; 35th WMA General Assembly, Venice, Italy, October 1983; 41st WMA General Assembly, Hong Kong, September 1989; 48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996; and 52nd WMA General Assembly, Edinburgh, Scotland, October 2000. Notes of Clarification added by the WMA General Assembly, Washington 2002 and Toyko 2004). Ferney-Voltaire, France: WMA. Retrieved April 13, 2010, from http://ohsr.od.nih.gov/guidelines/helsinki.html

Internet Links

[27] Afrobarometer Survey. Retrieved July 1, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[28] Asian Barometer. Retrieved July 1, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

[29] European Social Survey. Retrieved July 1, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=23&Itemid=318

[30] International Social Survey Programme. Retrieved July 1, 2010, from http://www.issp.org/

[31] Survey of Health, Ageing and Retirement in Europe. Retrieved July 1, 2010, from http://www.share-project.org

Further Reading

American Anthropological Association (AAA). (1998). *Code of ethics of the American Anthropological Association*. Arlington, VA: AAA. Retrieved May 13, 2008, from http://www.aaanet.org/committees/ethics/ethcode.htm

American Association for Public Opinion Research (AAPOR). (2006-07). *Best practices for survey and public opinion research*. Lenexa, KS: AAPOR. Retrieved May 13, 2008, from http://www.aapor.org/bestpractices

American Association for Public Opinion Research (AAPOR). (2005). *AAPOR code of professional ethics & practices*. Lenexa, KS: AAPOR. Retrieved May 13, 2008, from http://www.aapor.org/aaporcodeofethics

American Statistical Association (ASA). (1999). *Ethical guidelines for statistical practice*. Alexandria, VA: ASA. Retrieved May 5, 2010, from http://www.amstat.org/about/ethicalguidelines.cfm

Australian Market and Social Research Society (AMSRS). (2003). *Code of professional behaviour*. New South Wales, Australia: AMSRS. Retrieved September 7, 2008, from http://www.mrsa.com.au/index.cfm?a=detail&id=2649&eid=138

Bulmer, M. (1998). The problem of exporting social survey research. *American Behavioral Scientist, 42,* 153-167.

Council for International Organizations of Medical Sciences (CIOMS). (2002). *International ethical guidelines for biomedical research involving human subjects*. Geneva: CIOMS. Retrieved May 13, 2008, from http://www.cioms.ch/guidelines_nov_2002_blurb.htm

Council of American Survey Research Organizations (CASRO). (1997-2008). *CASRO code of standards and ethics for survey research*. Port Jefferson, NY: CASRO. Retrieved November 13, 2009, from http://www.casro.org/codeofstandards.cfm

Couper, M. P., & de Leeuw, E. D. (2003). Nonresponse in cross-cultural and cross-national surveys. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 155-177). Hoboken, NJ: John Wiley & Sons.

Dawson, L., & Kass, N. E. (2005). View of US researchers about informed consent in international collaborative research. *Social Science & Medicine, 61*, 1211-1222.

European Union (EU). (1995). *Directive 95/46/EC of the European Parliament and the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*. *Official Journal of the European Communities, 281*, 31-50. Retrieved April 14, 2010, from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HT

ML

Freed-Taylor, M. (1994). Ethical considerations in European cross-national research. *International Social Science Journal*, *142*, 523-532. Retrieved May 13, 2008, from http://www.unesco.org/most/ethissj.htm

Hershfield, A. F., Rohling, N. G., Kerr, G. B., & Hursh-Cesar, G. (1983). Fieldwork in rural areas. In M. Bulmer & D. P. Warwick (Eds.), *Social research in developing countries* (pp. 241-252). New York, NY: John Wiley & Sons.

ICC/ESOMAR. (2007). *ICC/ESOMAR international code on market and social research practice*. Amsterdam, The Netherlands: ESOMAR. Retrieved May 13, 2008, from http://www.iccwbo.org/uploadedFiles/ICC/policy/marketing/Statements/ICCESOMAR_Code_English.pdf

International Statistical Institute (ISI). (2009). *International Statistical Institute declaration on professional ethics*. The Hague, The Netherlands: ISI. Retrieved March 1, 2010, from http://isi.cbs.nl/ethics0index.htm

Kearl, B., Gafsi, S., Mbilinyi, M., Nabila, J., & Norman, D. (Eds.). (1976). *Field data collection in the social sciences: Experiences in Africa and the Middle East*. New York, NY: Agricultural Development Council.

Marshall, P. A. (2001). The relevance of culture for informed consent in U.S.-funded international health research. In *Ethical and policy issues in international research: Clinical trials in developing countries, Vol. II: Commissioned papers and staff analysis* (Chapter C). Bethesda, MD: National Bioethics Advisory Commission. Retrieved June 9, 2008, from http://bioethics.georgetown.edu/nbac/clinical/Vol2.pdf

Mohler, P. Ph., Pennell, B-E., & Hubbard, F. (2008). Survey documentation: Toward professional knowledge management in sample surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology.* New York/London: Lawrence Erlbaum Associates.

Mohler, P. Ph., & Uher, R. (2003). Documenting comparative surveys for secondary analysis. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 311-327). Hoboken, NJ: John Wiley & Sons.

National Bioethics Advisory Committee (NBAC). (2001). *Ethical and policy issues in international research: Clinical trials in developing countries, Vol. 1: Report and recommendations of the National Bioethics Advisory Commission submitted to the President of the United States, April 18, 2001*. Bethesda, Maryland: NBAC. Retrieved May 13, 2008, from http://bioethics.georgetown.edu/nbac/pubs.html

National Commission for Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont Report: Ethical principles and* *guidelines for the protection of human subjects of research* (DHEW Publication No. (OS) 78-0012). Washington, DC: Government Printing Office. Retrieved May 13, 2008, from http://ohsr.od.nih.gov/guidelines/belmont.html

Office for Human Research Protections. (2008). *Institutional review board guidebook.* Retrieved May 13, 2008, from http://www.hhs.gov/ohrp/irb/irb_guidebook.htm

Organisation for Economic Co-operation and Development. (1981). *Guidelines on the protection of privacy and transborder flows of personal data*. Retrieved June 11, 2008, from http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00&&en-USS_01DBC.html

Sugarman, J., Popkin, B., Fortney, J., & Rivera, R. (2001). International perspectives on protecting human research subjects. In *Ethical and policy issues in international research: Clinical trials in developing countries, Vol. II: Commissioned Papers and Staff Analysis* (Chapter E). Bethesda, MD: National Bioethics Advisory Commission. Retrieved May 13, 2008, from http://bioethics.georgetown.edu/nbac/pubs.html

Twumasi, P. A. (2001). *Social research in rural communities* (2nd ed.). Accra, Ghana: Ghana Universities Press.

United States Department of Commerce. (2000). *Safe Harbor overview.* Washington, DC: United States Department of Commerce. Retrieved May 5, 2010, from http://www.export.gov/safeharbor/eg_main_018236.asp

United States Department of Health and Human Services Office for Human Research Protections. (2005). *US code of federal regulations, Title 45, Part 46*. Washington, DC: United States Department of Health and Human Services Office for Human Research Protections. Retrieved May 13, 2008, from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm#46.102

United States Department of Health and Human Services Office for Human Research Protections. (2008). *Office for Human Research Protections (OHRP): International issues*. Washington, DC: United States Department of Health and Human Services. Retrieved May 13, 2008, from http://www.hhs.gov/ohrp/international/

United States Government Printing Office. (1949-1953). *Trials of war criminals before the Nuremberg Military Tribunals under Control Council Law No. 10*. Nuremberg, October 1946 - April 1949 (The Nuremberg Code). Washington, DC: U.S. Government Printing Office. Retrieved May 13, 2008, from http://www.hhs.gov/ohrp/references/nurcode.htm

Warwick, D. P. (1983). The politics and ethics of field research. In M. Bulmer & D. P. Warwick (Eds.), *Social research in developing countries* (pp. 315-

330). New York, NY: John Wiley & Sons.

World Association for Public Opinion Research (WAPOR). (n.d.). *WAPOR code of professional ethics and practices*. Lincoln, NE: WAPOR. Retrieved May 5, 2010, from http://www.unl.edu/wapor/ethics.html

World Health Organization (WHO). (2000). *Operational guidelines for ethics committees that review biomedical research* (Document TDR/PRD/ETHICS/2000.I). Geneva: WHO. Retrieved May 13, 2008, from http://www.who.int/tdr/publications/publications/ethics.htm

World Psychiatric Association (WPA). (2008). *Declaration of Madrid* (adopted WPA General Assembly, Madrid, Spain, August 1996; and enhanced by the WPA General Assemblies in Hamburg, Germany on August 8, 1999, in Yokohama, Japan, on August 26, 2002, and in Cairo, Egypt, on September 12, 2005). Geneva: WPA. Retrieved May 5, 2010, from http://www.wpanet.org/v1/content/madrid-ethic-engish.shtml

IV. Tenders, Bids, and Contracts

Rachel A. Orlowski and Christopher Antoun

Introduction

The following describes the process for preparing tenders, soliciting bids, and drawing up and executing contracts. The tenders should be based on the specifications outlined in the Study, Organizational, and Operational Structure chapter. Bids should be obtained from as many qualified organizations as possible to ensure a high quality survey for the budget available. The aim of signing contracts is to create legally-binding agreements between the organization coordinating the study and the organizations collecting the data.

Complications will inevitably arise over the course of the study, so it is important that the signed contracts define, in as much detail as possible, the specifications and expectations regarding procedures, responsible parties, and outcomes for all parts of the study across all participating organizations. Towards this end**,** the tenders, bids, and contracts should all be as specific and detailed as possible [11].

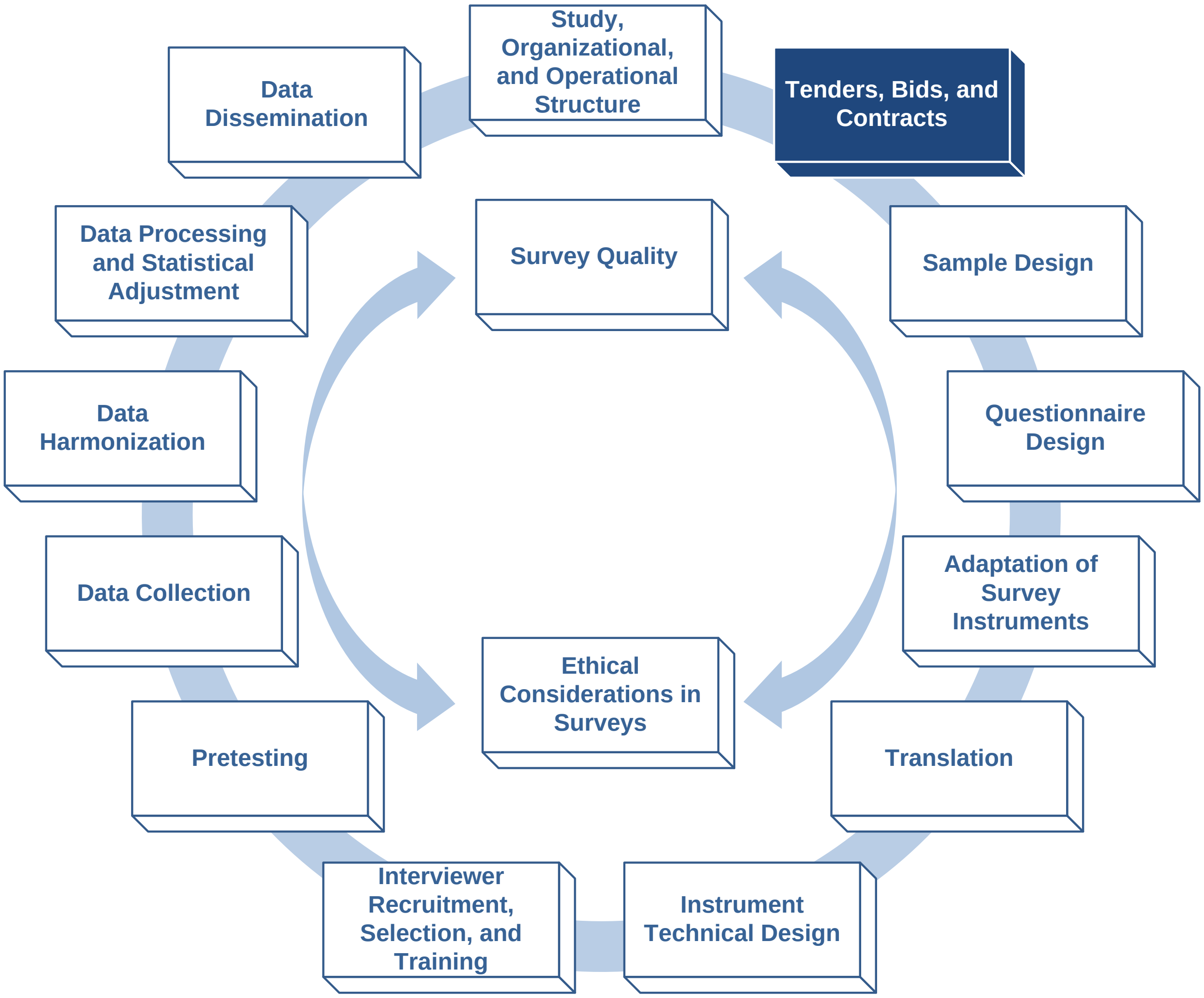
In cross-cultural studies, the tendering, bidding, and contracting process will involve various parties (e.g., the survey organizations, central coordinating center, and funder(s)), and there are many ways for these parties to conduct the process. For example, the coordinating center may prepare tenders and solicits bids from survey organizations in each country where the study will be carried out. The resulting contracts are between the coordinating center and each selected survey organization. There are other situations in which the coordinating center signs a contract with an international organization that is responsible for data collection in several countries. Sometimes the coordinating center submits a bid, rather than soliciting bids. In this case, the funder(s) prepares tenders for the central coordinating center and survey organizations separately. Thus, there is a contract between the funder(s) and the coordinating center, as well as separate contracts between the funder(s) and local survey organizations. Finally, there are other situations in which the central coordinating center is not involved with contractual work, and contracts are individually arranged and signed at the country level. In this situation, the central coordinating center may provide specifications and supervise the process, but the contract, however, is an agreement between the local funder(s) and local survey organizations.

The guidelines presented here deal with the more general approach to the tendering, biding, and contracting process described in the first case—outlining a competitive bidding process between a central coordinating center and survey organizations (in particular, survey organizations selected at the country level).

Nevertheless, many of the guidelines below also apply to the other contracting arrangements.

Figure 1 shows tenders, bids, and contracts within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To prepare tenders and conduct a competitive bidding process that will result in detailed contracts within defined budget parameters and an established legal framework.

1. **Prepare a tender based on study specifications while adapting it, when appropriate, for each individual country.**

Rationale

The tender is the first step to soliciting bids and executing contracts. The specifications in the tender will have long-term effects on the final contracts and the implementation of the study. With the tendering process, the coordinating center should consider the amount of risk it is willing to assume and specify the type of contract it will offer. The tender should outline study details and requirements of the bidding survey organizations. Requesting detailed information on technical and business aspects of the survey organization's bid reduces the opportunity for misunderstanding to go unnoticed and helps ensure that the study specifications have been fully understood and adequately accounted for in the plan and budget. In the final preparation of the tender, local adaptations should be considered, and multiple tenders may need to be developed for the many countries to set reasonable expectations based on the culture and availability of resources.

Procedural steps

- Determine the appropriate tendering process in each participating country.
    - Decide between open tendering and restricted tendering.
        - Open tendering allows any survey organization to provide a bid. It is advantageous because it protects against favoritism. Open tendering is absolutely necessary if the coordinating center is not familiar with the availability of qualified survey organizations in a country.
        - Restricted tendering limits the bidding process to a few survey organizations pre-selected by the coordinating center. Restricted tendering is used when the coordinating center has prior knowledge of survey organizations that are capable of implementing their country's portion of a cross-cultural study.
    - Become familiar with the local requirements for tendering (e.g., some countries prohibit restricted tendering if using public funds).

- State in the tender which type of contract will be offered: fixed-price, cost-reimbursable, or time and material. The decision on which type of contract will be offered depends on the level of risk the coordinating center (or funding source) and survey organizations are willing to take [9].
  - A fixed-price (or lump-sum) contract requires stating upfront a fixed total price for the clearly-defined scope of work and deliverable(s). Fixed-price contracts also allow for bonuses if expectations are exceeded. The coordinating center incurs little risk while the survey organizations incur much risk.
  - A cost-reimbursable contract requires paying the survey organizations for the actual costs necessary to complete the agreed-upon scope of work and production of the deliverable(s); it may include paying them a fee — typically received as profit. Cost-reimbursable contracts also allow for bonuses if expectations are exceeded. These guidelines assume cost-reimbursable contracts. This type of agreement is riskier for the coordinating center than for the survey organizations. Thus, it is important for the coordinating center to carefully evaluate survey organizations during the bidding process and to monitor progress during survey design and implementation.
  - A time and material (T&M) contract has elements of both the fixed-price and the cost-reimbursable contract. Time and material contracts may require a fixed level of effort by a specific class(es) of resources at the survey organizations or may have a variable level of effort by a specific class(es) of resources at an agreed-upon rate of pay for the specific class(es). These contracts may be open-ended, such that the exact price for the scope of work and/or deliverable(s) may not be determined when signing the contract. This type of contract is rarely used for the implementation of an entire survey project; it is sometimes used when contracting work for a particular task in the survey lifecycle (e.g., contracting with an organization to perform the post-collection data analysis).

- Ask bidders to provide specific technical information about their survey organization and their plan to execute the survey within the study specifications.
  - Request the following from the survey organization:
    - Examples of similar studies the bidder has conducted (describing the size, complexity, topic, etc.).
    - Examples of the bidder's training and supervisory materials, details of procedures used, and example reports from studies previously conducted.
    - References or contact names regarding previously completed work.

- Number and relevant qualifications of all levels of staff assigned to the study (also providing an organizational chart and outline of responsibilities for this survey).
- Organizational capacity (e.g., size of field interviewing staff).
- Technical system capability (e.g., any computer-assisted interviewing, sample management capabilities, or data entry software).
- Facilities and equipment (e.g., computers, internet access, and e-mail).

▪ Request the following regarding their plan to execute the survey:
- Timeline with survey tasks, milestones, and deliverables (see Study, Organizational, and Operational Structure for details about creating appropriate timelines and see Appendix A for an example of a timeline of effort).
- Staff responsibilities for each survey task (see Appendix B for an example of a person loading chart describing how responsibilities are assigned).
- Consent, confidentiality, and data protection procedures (see Ethical Considerations in Surveys).
- Sampling methods (e.g., sample size, type of frame, etc.) (see Sample Design) [2].
- Questionnaire development and translation methods (if needed).
- Pretesting methods (making sure, in addition to pretesting the questionnaire, that there is a pretest of the field procedures) (see Pretesting).
- Design of survey instrument (see Instrument Technical Design).
- Interviewer recruitment, selection and training protocol (number of hours of training, topics covered, etc.) (see Interviewer Recruitment, Selection, and Training).
- Interviewer characteristics (e.g., age, education, gender, and experience) (see Interviewer Recruitment, Selection, and Training).
- Unique identification of the interviewers (especially when recording which interviewers contacted which sample elements in the sample management system).
- Interviewer payment plan (typically by the hour or by completed interview) (see Interviewer Recruitment, Selection, and Training).
- Interviewer employment conditions (i.e., employees of the survey organization or contract workers).
- Ratio of interviewers to supervisors (see Data Collection).
- Mode of data collection (if using a mixed mode design, whether multiple modes will occur concurrently or sequentially) (see Data Collection).

        - Information about the contact attempts (e.g., time, day, interim disposition codes).
        - Production requirements (e.g., minimum number of contacts to attempt to obtain a complete interview, minimum response rate, etc.).
        - Local quality monitoring (e.g., evaluating recorded interviews, re-interviews on key survey items).
        - Plans in place to address nonresponse bias (see Data Collection).
        - Procedures for local data return (e.g., mail or electronic transfer of completed interviews and other materials).
        - Procedures for processing, managing, and storing data (see Data Processing and Statistical Adjustment).
        - Procedures for providing data to coordinating center.
        - Procedures developed to handle unexpected problems (i.e., risk management) [2].

- Ask bidders to organize the business information by each major survey task—tailoring the budget to the specific country's implementation of the study [3]. All of the chapters of the Cross-Cultural Survey Guidelines could be considered as viable survey tasks (see Appendix A in Study, Organizational, and Operational Structure for a brief description of each survey task).
    - Within each organized survey task, ask bidders to prepare a detailed budget by the two general categories: direct costs and indirect costs [9]. Direct costs typically consist of salary and non-salary costs, and indirect costs are typically calculated as a percentage of the total direct costs (both salary and non-salary costs).
        - Salary costs include: labor (both regular and temporary staff), fringe (calculated as a percentage of the regular staff labor costs), and overhead (calculated as a percentage of the total labor and fringe costs) [5]. For each staff position, budget the number of hours needed for each staff member for each survey task in which he or she will contribute. (See Appendix C for a salary budget example template that specifies labor hours for the pretesting task.)
        - Nonsalary costs include general sample purchase; supplies (e.g., pencils, folders, binders, etc.); printing (e.g., letterhead, training materials, respondent booklets, maps, reports, etc.); postage; communications (e.g., local and long distance telephone connect, high-speed internet connection, etc.); computing (e.g., laptop computers, printers, equipment maintenance, software licensing, security protection, etc.); interviewer recruitment (e.g., advertisements, community

meetings, etc.); hosting training (e.g., hotel arrangements, meals, etc.); travel (e.g., mileage, vehicle rental, vehicle maintenance, fuel, etc.); respondent incentives; and consultant fees (e.g., stipend, per diem, travel, etc.). (See Appendix D for a non-salary budget example template that specifies costs for the pretesting task.)

- Require bidders to write justifications for all direct and indirect costs, as well as to be explicit with the budgeting assumptions taken (e.g., the duration/dates of each survey task, the questionnaire length, the number of hours needed to receive a complete interview, the average distance interviews will travel, the expected response rate, the expected interviewer attrition rate, the cost of each supply item, etc.) [5].
- For study designs with a lot of uncertainty, advise bidders to include contingency (possibly 10%) into the budget to account for this risk [6].
- For studies lasting longer than one year, suggest the inclusion of a cost-of-living increase [5].
- For areas with rampant inflation, require frequent updates to the projected budget.

Lessons learned

- Contacting survey organizations upfront to discuss project details can help avoid possible complications during the bidding process, especially if a culture is unfamiliar with a formal bidding process.

- Gathering information about constraints on survey organizations before issuing tenders will improve the bidding process. These constraints include legal requirements, cultural norms, lack of organizational capacity (e.g., does not have computer-assisted interviewing capability), standard organizational practice (e.g., organization usually only provides interviewers two days of training but the tender requires a week), attitudes (e.g., different attitudes about data collection), etc. This information should be used to adapt specifications in tenders to each country as appropriate.

- Survey organizations may hesitate to mention any obstacles to conducting the study as outlined in the tender specifications for various reasons. Organizations should be encouraged in a culturally appropriate fashion to be open and explicit about anything that would conflict with the study specifications. Some obstacles may be quickly remedied if identified in advance. For example, it may be necessary to appoint male interviewers to some locations (such as lumber camps or mines) or to notify gatekeepers of the study and explain the need to contact given respondents. Strategies and schedules should be developed to accommodate this.

2. **Ensure a fair and competitive bidding process.**

Rationale

If the research capacity of a country is not already apparent to the central coordinating center, the bidding process is one way to illuminate this and to determine if any methodological or substantive expertise may be needed to supplement local resources. A competitive bidding process is not always possible; sometimes, there are only one or two competent survey organizations within each location being studied.

Procedural steps

- Request bids in a language understood by the reviewers from the central coordinating center, or arrange for language resources for the reviewing team to enable them to evaluate the bids.

- Provide bidders with the evaluation criteria, such that they will then know what is expected at each phase of the survey lifecycle as well as what deliverables are required at each phase (see Guideline 3) [3].

- Encourage consortium bids as seems relevant because, in contexts with sparse resources, partnerships may enable survey organizations to make stronger bids if together they have a broader set of proficiencies [10].

- Set a timeline for the bidding process that still allows time for clarification and discussion between the contracting parties and for several iterations of bids.

- Encourage bidding organizations to identify any elements required in the tender specifications that they are unable or unwilling to meet [2]. Doing so helps avoid bids which the bidding organizations will not fulfill.

- Check bids for potential problems, such as the following:
  - Can a proper sampling frame be obtained (see Sample Design and Data Collection)?
  - Does the bidding survey organization have access to the sample elements on the frame (e.g., will political conflicts or travel restrictions limit the areas in which the survey organization can contact individuals) (see Sample Design and Data Collection)?

- Is the concept of probability sampling understood and its implementation assured (see Sample Design)?
- Are suitable protocols and trainers available for interviewer training and interviewer motivation (see Interviewer Recruitment, Selection, and Training)?
- Are essential nonresponse bias reduction techniques realized (see Data Collection)?
- Are adequate quality control procedures in place (see Survey Quality)?
- Are necessary facilities, such as hardware, software, and internet access, available?
- Is the specification of budget details adequate?
- Are there local research "traditions," such as quota sampling or undocumented substitution, that may conflict with study specifications?

- Keep the bidding process transparent, open, and fair.
  - Provide the same level of help or assistance to every survey organization [4].
  - If new information becomes available that would be useful in preparing a bid, take care to distribute this information to all bidders.

Lessons learned

- Following up with the survey organizations to make sure they know what is expected is one way to maintain a fair bidding process. By clarifying aspects of the survey organization's bid, the coordinating center can avoid possible complications later in the implementation of the survey. For example, in many countries the research tradition is to pay interviewers by the completed interview and not by hours worked. The coordinating center may want to explain that this practice might work well if all interviewer assignments are of the same difficulty and if the length of the interview administration is within well-defined limits. However, if assignments vary in difficulty (longer travel times, for example) or the length of the interview can vary widely (dependent upon the respondent's answers), this will not work as well. It is important for the coordinating center to emphasize the endured risk of paying interviewers by the completed interview. Interviewers might be tempted to use strategies to keep interviews as short as possible in order to complete more cases. In the worst scenario, interviewers might be tempted to falsify the interview (i.e., interviewer falsification) (see Interviewer Recruitment, Selection, and Training and Ethical Considerations in Surveys).

3. Select the survey organization best suited to carry out the survey in each country within the constraints.

Rationale

The decision to select a survey organization or collaboration of organizations that will carry out the study, based on pre-specified and agreed-upon evaluation criteria, is a critical one. A poor choice of an organization will divert attention and resources away from other aspects of the study and may have a lasting impact on the entire endeavor.

Procedural steps

- Form a bid evaluation team within the coordinating center that is comprised of a substantive expert, a statistical advisor, a methodological advisor, and, as relevant, legal and local expertise.
  - When necessary, involve additional consultants throughout the contracting process (i.e., from preparing the tender to signing the contract) [4].
  - Ensure there are no pre-existing relationships between the bid evaluation team members and the bidding survey organizations.
  - Determine in advance the process for final decisions on survey organization selection, in case disagreements among the review team should arise.
  - Have each member evaluate the survey organizations individually and make written notes.
  - Organize among the team a group discussion of the strengths and weaknesses of various bids.
  - Even if there is only one bid for a given country, conduct evaluation as described above with notes and a group discussion.
  - If the final required work scope and budget cannot be met by the bidding organization(s), decide whether a new round of bids is necessary or if some other alternative is available.

- Use the following indicators as the basis of evaluation criteria for choosing an organization:
  - Local knowledge of the population of interest [11].
  - Organizational and staff expertise in the subject area and survey methods envisioned [4].
  - Knowledge of and experience with conducting similar types of survey (both the organization as a whole and the management/personnel assigned to the project) [2] [10].
  - Ability to estimate the costs to complete the entire work scope.
  - Transparency of procedures.

- Organization of field staff, including the planned supervisory structure and implementation strategy (e.g., whether interviewers are stationed throughout study areas or travel extensively in teams to different sampled locations).
- Demonstrated or projected ability to meet the timeline and various specified outcomes [10].
- Demonstrated or projected availability of management staff and statistical support.
- Affiliations with professional organizations.
- Cost.
- Methodological rigor and quality of the technical proposal.

- Find out as much about the culture as possible before negotiating strategies with survey organizations. In particular:
  - Make use of local or regional feedback about the survey organizations. It can be very useful to ask local contacts (these may not be directly local but at least in the region) to provide information about the organizations.
  - Try to become aware of any local tendencies in terms of management and likelihood of acknowledging obstacles. Encourage people to point out difficulties in terms of the knowledge of local tendencies. If you lack knowledge of what could be involved and do not have someone suitable to act as an informant, introduce the topics you need to know about (for example, “We have sometimes found organizations fear their bid will not be considered if they admit they have trouble meeting requirements. We have learned to recognize information about local constraints as very important. Is there anything you would like to raise with us?”).
  - Learn to wait longer than you may be accustomed for a response and listen attentively for indirect mention of a constraint.
  - Try to become aware of local survey traditions or their absence. If through preparation for local negotiations it becomes clear that the study specifications run counter to local traditions, ask for information about how the organization intends to address this difference [7].
  - Try to become adept at recognizing and addressing hesitancy, as people or organizations may be reluctant to engage in unfamiliar procedures.
  - If something is known or found to be unusual in a given context, ask for a demonstration of its usefulness.
  - Aim to persuade those involved to try out suggested techniques or help adapt them to local conditions before deciding on their use. In other words, avoid determining the feasibility of techniques before trying them out.

- Negotiate work scope and costs with the most promising organization.
  - If the specifications change significantly, then reopen the bidding process to all competitive organizations [4].
  - Agree upon alternative designs prior to signing the contract, since change is more difficult once a study has started [11].

- Throughout this selection process, do not rely on the same person to act as both translator and negotiator with the survey organizations.

- Notify unsuccessful bidders of your selection once the contract has been awarded. Supply them with your reasoning for selection, and provide feedback as to how they could be more successful in future bidding processes [4].

Lessons learned

- When evaluating survey organizations, one of the most difficult decisions made is determining whether a survey organization is truly capable of implementing what has been promised in its bid. If two competing survey organizations propose similar technical bids, it is not always prudent to select the organization with the less expensive business bid (even though not doing so might conflict with pre-determined bidding evaluation criteria). It is important to balance the proposed technical aspects (and timeline) and budget with the survey organization's (and staff's) experience and references. Prior work is often very foretelling of future work.

- When evaluating the proposed data collection timeline of each survey organization, seasonal effects must also be taken into account. One country's harvest time may be another's winter months; access to areas may be restricted or facilitated by the season. In certain times of year, large parts of the population may be on vacation or working away and difficult to reach at their usual residence [11].

4. **Execute a contract that addresses the rights and obligations of all parties involved and references local legal requirements, if applicable.**

Rationale

The final contract that the coordinating center drafts is legally binding and thus must fall under the auspices of a recognized legal authority with the power to sanction contract breaches. The sanctions should be explicit, up to and including nullifying the contract. The contract needs to be properly signed and dated by authorized representatives. Local, independent legal advice is critical to this process.

Procedural steps

- Write the contract based upon the study design and specifications.

- Tailor contracts to the funding source, contracting organizations, and countries, as necessary. Each may carry additional requirements, such as stipulated delivery of reports, a cross-national ethics board review, and so forth.

- Require official pre-approval of any subcontracting. Any known need for subcontracting in any form should be disclosed in advance by the survey organization(s) [4].

- Incorporate bonus schemes in the contract and cost estimates as appropriate. Examples may include:
  - Interviewer bonuses, based on performance.
  - Organizational bonuses, such as a payment for completing interviews beyond the expected total.

- Identify and specify the coordinating center's right to observe aspects of data collection (e.g., live interviews, call-backs to selected households for verification, spot checks of original questionnaires, and electronic control files) [8].

- Set reasonable production benchmarks, where possible [4].
  - Define targeted response rates as one of the production benchmarks (see Data Collection).
    - For the purpose of response rate calculation, provide the survey organizations with a defined list of the disposition codes to be used uniformly (see Appendices E - H for a description of disposition codes and templates for calculating response rates).
    - Go through the list of disposition codes, checking applicability of each for the local situation and define the need for additional codes to account for local conditions.
  - Require field monitoring progress reports (possibly at the individual interviewer, interviewing team, or region level) to ensure benchmarks are met (see Data Collection).

- Establish and specify in writing ownership of the data and respondents' sample and contact information within the limits of any confidentiality restrictions.

- Specify requirements for how the local survey organization will execute the data delivery and frequency of updates on data collection progress (see Data Collection, Data Harmonization, and Data Dissemination).

- Specify any deliverables (such as sample specifications, instrument specifications, and source questionnaires) and commitments from other parties involved, including any central organization to local organizations (e.g., advisory boards and help lines).
  - Identify and specify all required documents.
    - Agree on format for these and on who develops the format.
    - Include provisions for training for those required to provide documentation.
    - Consider requiring copies of the consent form, translated questionnaire, training materials, and methods report (see Ethical Considerations in Surveys and Interviewer Recruitment, Selection, and Training).

- Specify copyrights for data and documents, including stipulations for data release (by when and by whom) and plans for data access rights (taking into account any legal restrictions).

- Define the necessary security level of respondent data (e.g., contact information and survey responses) for both physical and electronic storage and transfer.

- Define any restrictions on the survey organization's ability to present and publish any of the substantive or methodological results (with or without review).

- For longitudinal studies, indicate, as appropriate, decisions about the protocol for possible respondent recontact. If potential for future follow-up exists, consider introducing this possibility at the time of initial contact with the respondents and ask the survey organization to budget for this activity.

- Instruct the survey organization to notify the coordinating center of any change to the contract [11].

Lessons learned

- Although it is important to enforce adherence to specifications, a measure of flexibility is also needed. Natural disasters, unexpected political events, and outbreaks of disease can interrupt data collection and make agreed-upon deadlines impossible (see Data Collection).

- Approving the use of subcontractors may impact the coordinating center's level of control. For example, monitoring data collection will be problematic if subcontractors restrict the right of the coordinating center to observe aspects of the survey process. Certain study specifications, such as the required security level of respondent data, could be difficult to ensure while working with subcontractors.

5. **Define upfront the quality standards that lay the quality framework for the rest of the survey lifecycle.**

Rationale

The bidding process may be the first interaction the survey organizations have with the coordinating center. Hence, it is essential for the coordinating center, from the conception of the survey, to demonstrate and emphasize the importance of quality.

Procedural steps

- Develop a quality management plan (see Survey Quality). Use this plan as the outline for expectations of the survey organizations throughout the entire study.

- Ask bidding survey organizations to detail their quality control and quality assurance procedures, and include minimum quality requirements in the criteria used for evaluating the bidders.

- Consider re-releasing the tender if no bidding survey organization can meet the requested quality standards.

- Define progress approval points throughout the research process (e.g., sample selection, questionnaire design, interviewer training, and data collection milestones) to ensure each party involved achieves the study's objectives.
  - Require certification from the coordinating center at these formal points before a survey organization can proceed with the study.
  - Sanctions for unnecessary delays or specification deviations should be specified before the study begins.

Lessons learned

- Since budgets are often underestimated, it is critical to monitor the overall budget throughout the survey lifecycle to avoid a potential overrun at the end of the study. In addition, individually monitoring the budget of each survey task is an important quality assurance procedure. If the budget for each survey task is more detailed (i.e., specified budgets for each direct cost component), it is useful to systematically assess the status of the budget and weigh the quality trade-off by monitoring costs at the lower levels (see Survey Quality).

6. **Document the steps taken while preparing tenders, soliciting bids, and drawing up and executing contracts.**

Rationale

The coordinating center can use the contract resulting from the bidding process to enforce its expectations of the survey organizations. Thus, it is very important that steps taken throughout the process be clearly noted and transparent to those involved. No one involved should be surprised at how the study is to be structured, what production actions are required, and when the final deliverables are to be completed.

Procedural steps

- Clearly state specifications in tenders.
- In advance of releasing tenders, document the evaluation criteria to be used when assessing bids.
- Keep a record of the information exchanged with each survey organization to make sure no one organization receives differential treatment during the bidding process.
- Document bid evaluation team scores for each survey organization's bid.
- Collect notes from each member of the bid evaluation team as to how they arrived at their selection decision.
- Make sure each survey organization formally details all aspects of their anticipated scope of work in their bid. Information beyond what is written in the bid (e.g., from other forms of correspondence) should not be considered when evaluating the survey organization—so as not to give differential treatment.
- Keep records of all notifications to unsuccessful bidders of your selection.

- Write contracts that are tailored to the involved parties (e.g., funding source, coordinating center, survey organization, etc.). When writing the contract, include all specifications of the scope of work, budget, and timeline for which each survey organization should commit.

- In the contract, establish responsibility for documenting all aspects of the study.

- Request documentation of any subcontracts from the survey organizations.

- Have a signed agreement regarding the ownership of the data and respondent information, within the limits of confidentiality regulations (see Ethical Considerations in Surveys).

- Keep a copy of the tenders, all bid materials provided by any survey organization submitting a bid, and a copy of the contracts (and any modifications).

Appendix A

Timeline of effort by survey task example

When bidding for a cross-cultural survey, it is important for a survey organization to outline how it plans to meet the specified deliverables' deadlines. This can be achieved by creating a timeline that demonstrates when the survey organization will work on each task of the survey lifecycle and how much effort (i.e., how many hours) is necessary to perform that task.

Below is an example of a timeline with an expected 24-month duration (specified in actual calendar months and years) and survey tasks corresponding with each chapter of the Cross-Cultural Survey Guidelines. The 'X's are placeholders for the number of hours assigned to each survey task per month (for the entire staff's effort). It is critical that the total number of hours for all tasks for all months equal the total number of hours for all assigned staff (see Appendix B).

TASK	MONTH OF SURVEY																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
Study, Organizational, and Operational	X	X																							
Survey Quality	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Ethical Considerations in Surveys	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Tenders, Bids, and Contracts	X	X	X																						
Sample Design		X	X	X	X	X																			
Questionnaire Design		X	X	X	X	X	X	X	X																
Adaptation				X	X	X	X	X	X																
Translation						X	X	X																	
Instrument Technical Design				X	X	X	X	X	X																
Interviewer Recruitment, Selection, and Training				X	X	X	X	X				X				X									
Pretesting						X	X	X																	
Data Collection									X	X	X	X	X	X	X	X	X	X							
Data Harmonization																			X	X	X				
Data Processing and Statistical Adjustment												X	X	X	X	X	X	X	X	X	X	X	X		
Data Dissemination			X					X					X					X					X	X	
TOTAL																									

Appendix B

Person loading chart example

When bidding for a cross-cultural survey, it is important for a survey organization to outline how it plans to assign responsibilities for each task of the survey lifecycle to which staff members and how much effort (i.e., how many hours) is necessary for that staff to accomplish each given task. This can be achieved by creating a person loading chart.

Below is an example of a person loading chart with example study roles and survey tasks corresponding with each chapter of the Cross-Cultural Survey Guidelines. If the name of the staff member fulfilling the role is known, include the name. If the name of the staff member is not known, include the job title. Indicate if multiple people will be necessary for a given role. The 'X's are placeholders for the number of hours budgeted to staff for each task. It is critical that the total number of hours for all staff, for all tasks, equal the total number of hours for all months of the survey (see Appendix A.)

(Person loading chart)

	SURVEY TASKS															
STAFF	**Study, Organizational, and Operational Structure**	**Survey Quality**	**Ethical Considerations in Surveys**	**Tenders, Bids, and Contracts**	**Sample Design**	**Questionnaire Design**	**Adaptation**	**Translation**	**Instrument Technical Design**	**Pretesting**	**Interviewer Recruitment, Selection, and Training**	**Data Collection**	**Data Harmonization**	**Data Processing and Statistical Adjustment**	**Data Dissemination**	**Total**
Project Manager	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Quality Coordinator	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Budget Analyst	X	X		X												
Office Assistant	X	X	X	X						X	X	X				
Statistician		X			X					X		X	X	X	X	
Questionnaire Designer		X				X	X	X	X	X						
Translator		X				X	X	X	X							
Data Manager		X			X				X	X		X	X	X	X	
Information Technologist		X							X	X		X	X	X	X	
Programmer		X							X	X		X		X		
Field Manager		X							X	X	X	X		X		
Field Support Staff		X							X	X		X				
Data Processing Manager		X				X								X		
Interviewer Recruiter		X									X					
Interviewer Supervisor		X								X	X	X				
Interviewer		X								X	X	X				
Interviewer Aide/Driver		X								X		X				

Appendix C

Salary budget template example

Specifying the salary costs for each survey task is an important component of a bid. For each staff member, estimate the number of hours that he or she will contribute. In this example, the staff members expected to work on the pretesting task are listed by job title, with only one person needed for each role.

If the name of the staff member completing the role is known, include the individual's name and actual hourly rate. If the name of the staff member is not known, include the job title and average hourly rate for the staff members with that title. If several people have the same job title, include separate entries for each (however, "interviewers" may be listed as a single line). When a survey task is completed across multiple years, the budget estimate should account for the expected changes in hourly rates.

STAFF	HOURS	HOURLY RATE	TOTAL COST
Project Manager			
Quality Coordinator			
Office Assistant			
Statistician			
Questionnaire Designer			
Data Manager			
Information Technologist			
Programmer			
Field Manager			
Field Support Staff			
Interviewer Supervisor/Trainer			
Interviewers			
Interviewer Aide/Driver			
TOTAL HOURS:		TOTAL COST:	

Appendix D

Non-salary budget template example

Specifying the non-salary costs for each survey task is an important component of a bid. For each survey task, estimate the number of items and cost per unit. In this example, the items expected to be used for the pretesting task are listed. When a task is completed across multiple years, the budget estimate should account for the increases in per unit material costs.

ITEMS	NUMBER	COST PER UNIT	TOTAL COST
Facilities			
Meals			
Laptop Computer			
Software Licensing			
General Supplies			
Communications			
Postage			
Printing			
Respondent Incentive Payments			
Respondent Recruitment Expenses			
Travel			
		TOTAL COST:	

Appendix E

Disposition codes [1]

- The coordinating center should provide a list of specific disposition codes with a clear description of how to code all sample elements during (temporary disposition) and at the close of (final disposition codes) the field period.

- Generally, disposition codes identify sample elements as (complete or partial) interviews or non-interviews.
  - The coordinating center should set the criteria for determining whether interviews are classified as complete or partial.
  - Non-interviews are grouped by whether the respondent is eligible, unknown eligible, or ineligible to participate in the study.

- Disposition codes are mutually exclusive. While sample elements may be assigned different temporary disposition codes throughout the field period, there will be only one final disposition code.

Appendix F

Components and descriptions of each category of response rate calculation (for a sampling frame of housing units) [1]

- To standardize the response rate calculations across countries, every country should group each sample element's final disposition code into one of the following mutually exclusive and exhaustive categories:
  A. Interviews
  B. Non-interviews—Eligible
  C. Non-interviews—Unknown eligibility
  D. Non-interviews—Ineligible

A. Interviews

Component	*Description*
Complete interviews	• Respondent has finished the interview.
Partial interviews	• The survey organization (in consultation with the coordinating center) may decide prior to the start of data collection to consider an interview to be a partial interview if at least some percent (e.g., 80%) of applicable or crucial/essential questions have been answered.
TOTAL INTERVIEWS	• Sum of interviews.

B. Non-interviews—Eligible

Component	*Description*
Refusals	• It has been determined that there is an eligible respondent in the housing unit but either he/she or someone else refuses the interview request.
Non-contacts	• It has been determined that there is an eligible respondent in the housing unit but the interviewer cannot gain access to the building, no one is reached at the housing unit, or the respondent is never available when the interviewer attempts an interview.
Other	• It has been determined that there is an eligible respondent in the household

(Other)	(eligibility determined as of a particular date, e.g., the date that the household listing is taken) but at some time after the determination of eligibility, the respondent is unable to complete the interview due to reasons other than a refusal or is unable to be reached after repeated attempts. For example, the respondent may have died, been incarcerated or hospitalized, or left the country. • It has been determined that there is an eligible respondent in the household, but he/she does not speak any of the study language(s) or is permanently incapable of participating in the interview due to a physical or mental condition (e.g., senility, blindness, or deafness). Note: Sample elements may be considered ineligible if the target population is defined such that respondents who do not speak the study language(s) or respondents who are unable to hear are explicitly excluded from the target population to which the study plans to makes inferences. • Any other eligible non-interview status.
TOTAL NON-INTERVIEWS—ELIGIBLE	• Sum of eligible non-interviews. • If the survey organization is unable to provide separate counts of each component but the survey organization can provide the total number of eligible non-interviews, use the total.

C. Non-interviews—Unknown eligibility

Component	*Description*
Unknown if household/occupied housing unit	• The sample elements have not been attempted or worked (e.g., no interviewer is available in area or replicates are introduced too late to work all sample elements). • Interviewer is unable to reach the

	housing unit due to weather or concerns about safety in a dangerous neighborhood. • Interviewer is unable to locate the housing unit (e.g., inaccurate or inadequate address/locating information).
Unknown if eligible respondent is in unit/no screener completed	• It has been determined that there is an eligible housing unit but the interviewer is unable to determine whether there is an eligible respondent in the unit. For example, a household member may refuse to complete the screener or no one is available to complete the screener when the interviewer visits the household. Note: These sample elements are not considered refusals, since only elements where it has been determined that there is an eligible respondent can be classified as refusals.
Other	• Any other status for which eligibility is unknown
TOTAL NON-INTERVIEWS—UNKNOWN ELIGIBILITY	• Sum of non-interviews of unknown eligibility • If the survey organization is unable to provide separate counts of each component, but the survey organization can provide the total number of non-interviews of unknown eligibility, use the total.

D. Non-interviews—Ineligible

Component	*Description*
Not an eligible housing unit	• The sample elements are out-of-sample housing units or housing units that are incorrectly listed in the address frame (e.g., housing units are outside the primary sampling unit in which they are thought to be located). • The sample elements are non-residential units (e.g., businesses, government offices, institutions, or group quarters).

(Not an eligible housing unit)	• Housing units are vacant on the date that eligibility is determined. Note: Sample elements may be considered eligible non-interviews if someone is present at the housing unit on the date that eligibility is determined, even if when the interviewer returns the household has moved and the unit is vacant. • Households are temporary, seasonal, or vacation residences (i.e., not the usual place of residence).
No eligible respondent	• It has been determined that there is an eligible housing unit, but there is no eligible respondent in the unit. For example: ▪ Residence with no one 18 years of age or older. ▪ Respondent does not speak any of the study language(s) and the target population is explicitly defined such that respondents who do not speak the study language(s) are not considered part of the target population to which the study plans to make inferences (may also hold for physical or mental conditions, if the target population is explicitly defined to exclude persons who are blind, deaf, senile, etc.). ▪ Respondent died before eligibility is determined. ▪ Respondent is incarcerated or hospitalized (i.e., institutionalized) at the time that eligibility is determined, and remains institutionalized throughout the data collection period.
Other	• Respondent is in a group/cell for which the quota has already been filled. • Any other ineligible non-interview status.
TOTAL NON-INTERVIEWS—INELIGIBLE	• Sum of ineligible non-interviews. • If the survey organization is unable to provide separate counts of each component but the survey organization can provide the total number of ineligible

	non-interviews, use the total.

Appendix G

Recording counts of response rate categories template
(for a sampling frame of housing units) [1]

- Use the template below to help determine the number (or weighted count, if appropriate) of sample elements finalized in each of the categories and, thus, the total number/weighted count of sample elements fielded. The total number of sample elements is the sum of all categories of the final disposition codes.
  - First, enter the number of sample elements finalized as each given category component. If no sample elements are finalized as a particular category component, enter “0” in the “Count” column.
  - Next, total the components for each category by entering the sum on the longer of the “Count” column lines.
  - Finally, total the sums of each category by entering the overall sum on the last “Count” column line.
  - Use the “Additional Information” column to provide any information that will assist in interpreting the figures provided, particularly the study’s definition of partial interviews or descriptions of “Other” classifications specific to the study.

Category (with Components)	*Count*	*Additional Information*
A. Interviews		
Complete interviews	___	___
Partial interviews	___	___
TOTAL INTERVIEWS	___	___
B. Non-interviews—Eligible		
Refusals	___	___
Non-contacts	___	___
Other	___	___
TOTAL NON-INTERVIEWS—ELIGIBLE	___	___
C. Non-interviews—Unknown eligibility		
Unknown if household/occupied housing unit	___	___
Unknown if eligible respondent in unit/no screener completed	___	___
Other	___	___
TOTAL NON-INTERVIEWS—UNKNOWN ELIGIBILITY	___	___
D. Non-interviews—Ineligibility		
Not an eligible housing unit	___	___
No eligible respondent	___	___
Other	___	___
TOTAL NON-INTERVIEWS—INELIGIBILITY	___	___
TOTAL NUMBER OF SAMPLE ELEMENTS	___	___

Appendix H

Recording counts of response rate categories for additional eligible respondents template
(for a sampling frame of housing units) [1]

- Use the template below to help determine the number (or weighted count, if appropriate) of additional respondents in each of the categories and, thus, the total number/weighted count of additional respondents. The total number of additional respondents is the sum of only the eligible categories of the final disposition codes; if a household was not eligible, no respondents—let alone additional respondents—were selected.
  - First, enter the number of additional respondents finalized as each given category component. If no additional respondents are finalized in a particular category component, enter "0" in the "Count" column.
  - Next, total the components for each category by entering the sum on the longer of the "Count" column lines.
  - Finally, total the sums of each category by entering the overall sum on the last "Count" column line.
  - Use "Additional Information" column to provide any information that will assist in interpreting the figures provided, particularly the study's definition of partial interviews or descriptions of the "Other" classification specific to the study.

Category (with Components)	*Count*	*Additional Information*
A. Interviews		
Complete interviews	___	___________________
Partial interviews	___	___________________
TOTAL INTERVIEWS	______	___________________
B. Non-interviews—Eligible		
Refusals	___	___________________
Non-contacts	___	___________________
Other	___	___________________
TOTAL NON-INTERVIEWS—ELIGIBLE	______	___________________
TOTAL NUMBER OF ADDITIONAL		

RESPONDENTS	______	____________________

Glossary

Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Bid	A complete proposal (submitted in competition with other bidders) to execute specified jobs within prescribed time and budget, and not exceeding a proposed amount.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Coding	Translating nonnumeric data into numeric fields.
Concurrent mixed mode	A mixed mode design in which one group of respondents uses one mode and another group of respondents uses another.
Confidentiality	Securing the identity of and any information provided by the respondent to ensure to the greatest extent possible that public identification of an individual participating in the study and/or his individual responses does not occur.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.

Coordinating center	A research center that facilitates and organizes cross-national research activities.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
Direct cost	An expense that can be traced directly to (or identified with) a specific cost center or is directly attributable to a cost object such as a department, process, or product.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Imputation	Computational methods that assign one or more estimated answers for each item that previously had missing, incomplete or implausible data.
Indirect cost	An expense that is incurred in joint usage and difficult to assign to or is not directly attributable to a specific department, process or product.

Interviewer falsification	Intentionally departing from the designed interviewer guidelines, which could result in the contamination of the data. Falsification includes: 1) Fabricating all or part of an interview – the recording of data that are not provided by a designated survey respondent and reporting them as answers of that respondent; 2) Deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a refusal case as ineligible for the sample; reporting a fictitious contact attempt); 3) Deliberately miscoding the answer to a question in order to avoid follow-up questions; 4) Deliberately interviewing a nonsampled person in order to reduce effort required to complete an interview; or intentionally misrepresenting the data collection process to the survey management.
Item nonresponse, item missing data	The lack of information on individual data items for a sample element where other data items were successfully obtained.
Longitudinal study	A study where elements are repeatedly measured over time.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Mode	Method of data collection.
Noncontact	Sampling units that were potentially eligible but could not be reached.
Non-interview	A sample element is selected, but an interview does not take place (for example, due to noncontact, refusal, or ineligibility).
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Nonresponse bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.
Open tendering	A bidding process in which all the bidders are evaluated and then chosen on the basis of cost and error tradeoffs.

Overrun The exceeding of costs estimated in a contract.

Post-survey adjustments Adjustments to reduce the impact of error on estimates.

Pretesting A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.

Primary Sampling Unit (PSU) A cluster of elements sampled at the first stage of selection.

Probability sampling A sampling method where each element on the sampling frame has a known, non-zero chance of selection.

Quality The degree to which product characteristics conform to requirements as agreed upon by producers and clients.

Quality assurance A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.

Quality audit The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.

Quality control A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.

Quality management plan A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.

Quota sampling	A non-probability sampling method that sets specific sample size quotas or target sample sizes for subclasses of the target population. The sample quotas are generally based on simple demographic characteristics (e.g., quotas for gender, age groups, and geographic region subclasses).
Recontact	To have someone other than the interviewer (often a supervisor) attempt to speak with the sample member after a screener or interview is conducted, in order to verify that it was completed according to the specified protocol.
Replicates	Systematic probability subsamples of the full sample.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Restricted tendering	A bidding process in which only bidders prequalified through a screening process may participate in bidding, in which they are evaluated and then chosen on the basis of cost and error tradeoffs.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sample management system	A computerized and/or paper-based system used to assign and monitor sample units and record documentation for sample records (e.g., time and outcome of each contact attempt).
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.

Sampling units Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.

Secondary Sampling Unit (SSU) A cluster of elements sampled at the second stage of selection.

Sequential mixed mode A mixed mode design in which additional modes are offered as part of a nonresponse follow-up program.

Source questionnaire The questionnaire taken as the text for translation.

Substitution A technique where each nonresponding sample element from the initial sample is replaced by another element of the target population, typically not an element selected in the initial sample. Substitution increases the nonresponse rate and most likely the nonresponse bias.

Survey lifecycle The lifecycle of a survey research study, from design to data dissemination.

Survey population The actual population from which the survey data are collected, given the restrictions from data collection operations.

Target population The finite population for which the survey sponsor wants to make inferences using the sample statistics.

Task	An activity or group of related activities that is part of a survey process, likely defined within a structured plan, and attempted within a specified period of time.
Tender	A formal offer specifying jobs within prescribed time and budget.
Total survey error	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

References

[1] American Association for Public Opinion Research (AAPOR). (2009). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (6th ed.). Retrieved April 2, 2010, from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814

[2] ESOMAR. (2001). *ESOMAR world research codes & guidelines: How to commission research* [Brochure]. Amsterdam: The Netherlands.

[3] Federal Committee on Statistical Methodology. (1983). *Statistical policy working paper 9: Contracting for surveys*. Washington, DC: Office of Management and Budget.

[4] Fink, A. (Ed.). (1995). *The survey kit: The survey handbook* (Vol. 1). Thousand Oaks, CA: Sage Publications.

[5] Glewwe, P. (2002). Overview of the implementation of household surveys in developing countries. In United Nations Statistical Division (Eds.), *Household sample surveys in developing and transition countries* (pp. 54-66). New York, NY: United Nations.

[6] Harik, I. (1987). Some political and cultural considerations bearing on survey research in the Arab world. In M. A. Tessler, M. Palmer, T. E. Farah, & B. L. Ibrahim (Eds.), *The evaluation and application of survey research in the Arab world* (pp. 66-74). Boulder, CO: Westview Press.

[7] International Organization for Standardization. (2006). *Market, opinion and social research—terms, definitions and service requirements* (ISO 20252). Geneva. Retrieved April 30, 2010, from https://www.astandis.at/shopV5/Preview.action%3Bjsessionid=0D8F503B3BD48375403D96257760FABB?preview=&dokkey=342276&selectedLocale=en/

[8] Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (Eds.). (2007). *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. London: Sage Publications.

[9] Project Management Institute. (2004). *A guide to the project management body of knowledge (PMBOK® guide)* (3rd ed.). Newtown Square, PA: Project Management Institute, Inc.

[10] United Nations. (2005). *Household surveys in developing and transition countries.* New York, NY: United Nations, Department of

Economic and Social Affairs. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[11] Worcester, R., Lagos, M., & Basañez, M. (2000). *Ccross-national studies: Lessons learned the hard way.* Paper presented at the 2000 AAPOR/WAPOR Conference, Portland, OR.

V. Sample Design

Frost Hubbard and Yuchieh Lin

Introduction

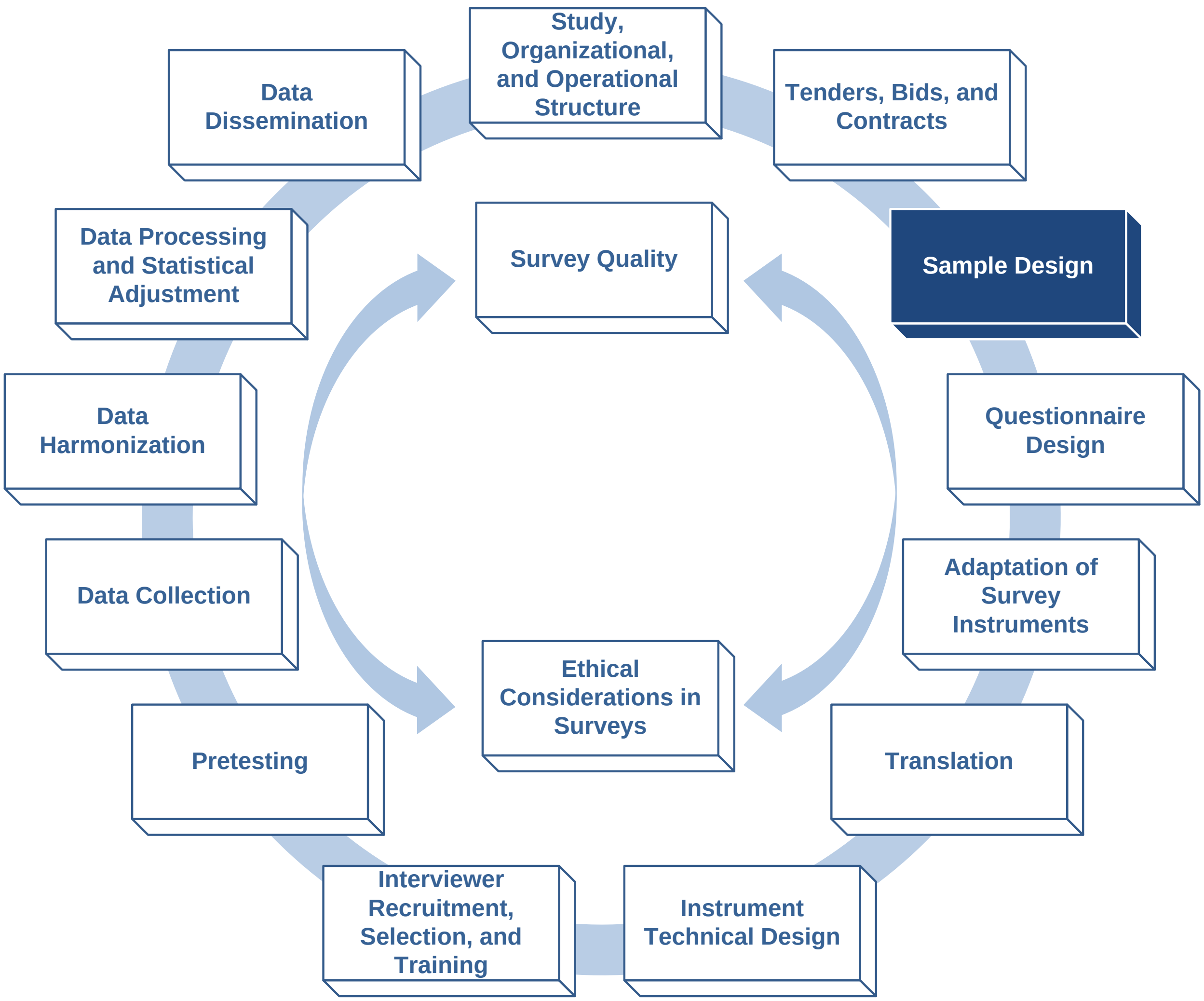
Optimal sample design can be defined as a probability sample design (see probability sampling) that maximizes the amount of information obtained per monetary unit spent within the allotted time and meets the specified level of precision [16]. One important prerequisite for comparative surveys is that all samples are full probability samples from comparable target populations [25]. Different nations have different sampling resources and conditions. For a cross-cultural survey, this means that the optimal sample design for one country may not be the optimal design for another. (Please note this chapter uses the term "participating country" to encompass any participating country, culture, region or organization in a cross-cultural study.) Therefore, allowing each participating country flexibility in its choice of sample design is highly recommended, so long as all sample designs use probability methods at each stage of selection [14] [25].

This chapter outlines the decisions that need to be made when designing a cross-cultural probability survey sample. It encourages cross-cultural survey organizers to allow sample designs to differ among participating countries while, at the same time, ensuring standardization on the principles of probability sampling.

Please note that this chapter assumes that the reader has a basic understanding of statistics and terms such as "variance" and "standard deviation." Please refer to Further Reading or an introductory statistics textbook if a statistics refresher is needed.

Figure 1 shows sample design within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To select an optimal, cost-efficient probability sample in each participating country that is representative of the target population and allows researchers to make inferences to the target population, and to standardize sample designs without hampering optimal designs in each participating country.

1. **Decide whether to administer a cross-sectional survey or one of the types of panel surveys.**

Rationale

Sometimes the decision regarding whether the sample survey should collect data from selected elements at only one point in time or at more than one point in time is clear-cut. In many situations, however, the decision is not straightforward, and survey organizers are wise to consider the benefits and drawbacks of each method. This decision will affect all aspects of the survey, including the cost, level of effort, and speed with which the results and analysis can be presented.

Procedural steps

- Consider the advantages and disadvantages of a cross-sectional survey (i.e., a survey where data are collected from selected elements at one point in time).
  - Advantages of cross-sectional surveys:
    - Since data are collected at only one point in time, countries can create an optimal sample design for that specific point in time.
    - Changes in the target population can be accommodated.
    - Since sampling units are only asked to participate once, the respondent burden over time is less than it would be in a panel survey; this can make it easier to convince the sampling units to participate.
  - Disadvantages of cross-sectional surveys:
    - Developments or changes on the individual level over time cannot be measured.

- Consider the advantages and disadvantages of a panel survey (i.e., a survey where the data are collected from selected elements at more than one point in time [3] [24] [28]). Panel surveys include fixed panel, fixed panel plus births, repeated panel, rotating panel, and split panel studies.
  - Advantages of panel surveys:
    - The ability to measure changes over time on the statistics of interest at the respondent level is greater.
  - Disadvantages of panel surveys:
    - The sampling design, being optimal at the outset of the panel survey, may be dated and not optimal at a later point in time.
    - Changes in the target population are difficult to implement (e.g., including non-citizens at a later stage).
    - The initial cost of a panel survey is higher than a cross-sectional survey since both thought and effort need to be expended to plan the best way to capture data over time.
    - It can be difficult to convince respondents to participate across multiple waves of data collection.

- With each successive wave of data collection, the cumulative amount of respondent attrition typically increases. Unless the element sample from the original wave of data collection is supplemented with fresh cohorts, the remaining respondents may not accurately reflect the target population.
- For surveys of mobile populations, the attrition rate can be very high. Survey planners should consider how to identify and track panel survey respondents, especially when dealing with a mobile population.
- Question wording and response options need to be comparable across waves in order to allow comparison over time on the statistic of interest.
- In contrast to a cross-sectional design, a comparative panel survey design implemented across many countries is much more complex. Designers should consider the efforts necessary to achieve comparability simultaneously across each national panel wave and across all countries.

Lessons learned

2. The World Fertility Survey (WFS), its successor, the Demographic and Health Survey (DHS), and the International Social Survey Programme (ISSP) are well-known cross-cultural studies which have demonstrated that large-scale probability sample surveys are feasible almost everywhere. For all participating countries in these two studies, sampling frames and resources (including households) were found; local technicians executed complex tasks directed by a centralized international staff; and probability sampling and measurable sampling errors were imposed [25] [34].

- Survey planners are not always aware of the time and effort required to design and implement quality cross-sectional sampling designs simultaneously across many countries. It might be instructive to consult the extensive documentation of the European Social Survey that includes design, control, and outcomes [46].

- Survey planners are sometimes naïve about the high cost and effort required to maintain a panel survey. When considering the implementation of a panel survey, refer to the literature on longitudinal survey programs such as the Survey of Income and Program Participation [19], the British Household Panel Survey [29], the European Community Household Panel [33], Canada's Survey of Labour and Income Dynamics [26], and additional literature about the methods used in longitudinal surveys [29] and panel surveys [20]. This literature gives a clear sense of the effort and expense necessary to execute a panel survey, and can help survey planners make a more judicious decision regarding the time dimension of the survey design.

2. **Define the target population for the study across all countries and the survey population within each participating country.**

Rationale

The survey planners of any cross-cultural survey need to develop a detailed, concise definition of the target population in order to ensure that each participating country collects data from the same population. Without a precise definition, one country may collect data that include a certain subgroup, such as noncitizens, while another country excludes this subgroup. This difference in sample composition may influence the estimates of key statistics across countries. In addition, a precise definition will let future users of the survey data know to which exact population the survey data refer. The data users can then make a more informed decision about whether to include the survey data in their analyses.

Procedural steps

- Define the target population across all participating countries as clearly as possible, including what units are elements of the populations and the time extents of the group [10]. For example, a target population might be defined as, “All persons above the age of eighteen who usually slept most nights in housing units in South Africa, Zimbabwe, Lesotho, and Swaziland during April, 2007.” (Note that this definition would, in turn, require definitions of the terms “usually,” “most,” and “housing unit.”)

- To ensure a clear description of the target population, think about all the potential inclusion/exclusion criteria. For example, the target population might exclude:
  - Persons outside a defined age range.
  - Persons in institutions, such as hospitals, nursing homes, prisons, group quarters, colleges, monasteries, or military bases.
  - Persons living in certain sparsely populated or remote geographic regions.
  - Non-citizens, ethnic minorities, homeless or nomadic populations, language groups.

- Define the survey population within each participating country by refining the target population based on cost, security, or access restrictions to all target population elements [10]. Make sure that the resulting survey populations are comparable across all countries.

Document for which region/strata the survey population allows inferences in later analyses.

- For example, the survey population may exclude those residing in war-torn areas, or the data collection period may be narrowed in areas with civil disturbances that are threatening to escalate.

Lessons learned

- Large established cross-cultural surveys have defined their target and survey populations differently, depending upon the goals and topics of the study.
  - The Afrobarometer is an independent, nonpartisan research project that measures the social, political, and economic atmosphere in Africa. Afrobarometer surveys are conducted in more than a dozen African countries and are repeated on a regular cycle. Participants in Round 4 of the Afrobarometer Survey had to be citizens of their country and of voting age the day of the survey. They had to complete the interview in their country's national language or in an official local language translation. Areas of armed conflict or natural disasters, national parks and game reserves, and people living in institutionalized settings were excluded. Special cases, like areas of political unrest, were reviewed on a case-by-case basis [44].
  - The Asian Barometer (ABS) is an applied research program studying public opinion on political values, democracy, and governance in thirteen East Asian political systems (Japan, Mongolia, South Koreas, Taiwan, Hong Kong, China, the Philippines, Thailand, Vietnam, Cambodia, Singapore, Indonesia, and Malaysia) and five South Asian countries (India, Pakistan, Bangladesh, Sri Lanka, and Nepal). The target population of the Asian Barometer was defined as citizens who were at least 20 years of age and were eligible to vote (i.e., were not disenfranchised due to mental illness or incarceration) [45].
  - The European Social Survey (ESS) is an academically-driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behavior patterns of its diverse populations. Round 5 of the ESS covers more than 30 nations and includes persons 15 years or older who are resident within private households, regardless of nationality, citizenship, or language; homeless and institutional populations are excluded from the sample [47].
  - The Living Standards Measurement Study (LSMS) was established by the World Bank in 1980 to explore ways of improving the type and quality of household data collected by statistical offices in developing countries. Its goal is to foster increased use of household data as a basis for policy decision making. Respondent

requirements and exclusions vary across participating countries [27].

- The Survey of Health, Ageing and Retirement in Europe (SHARE) studies the health, socio-economic status and social and family networks of individuals, aged 50 or over, in countries ranging from Scandinavia (Denmark, Sweden) through Central Europe (Austria, France, Germany, Switzerland, Belgium, the Netherlands, the Czech Republic, Poland) to the Mediterranean (Spain, Italy, Greece, Israel), as well as Ireland. In addition to the age requirement, respondents are residents and their partners (independent of partner age) who speak the official language. The study excludes seasonal or vacationing residents, persons physically or mentally unable to participate, those who died before the start of the field period, or who are unable to speak the specific language of the national questionnaire. It also excludes residents of institutions, except facilities for the elderly [49].
- The World Value Survey is conducted by a non-profit association seated in Stockholm, Sweden, to help social scientists and policy makers better understand worldviews and changes that are taking place in the beliefs, values, and motivations of people throughout the world. Respondents are adults, 18 years and older; some countries also place upper limits on age [42] [51].
- The World Mental Health (WMH) Survey studies mental illness in selected countries in Europe, Asia, Africa, and North and South America. One of the major goals of the WMH Study was to compare the age of onset of disease across countries. Best practice might suggest strictly defining the age of majority (e.g., 18 years old). However, the WMH study organizers recognized that strictly defining this inclusion criterion would be difficult, given that age of majority varies by country (and even within a country). Also, a strict definition would affect study protocols such as ethics reviews and informed consent (seeking permission to interview minors). Therefore, the WMH Study had to make a difficult decision about whether to strictly define the age eligibility criterion or allow it to vary across countries. In the end, the WMH Study allowed the age range to vary, with 16 years of age being the youngest lower age limit; some countries also set upper age limits. . This was taken into consideration in the analysis stage [50]. Participating countries were also allowed to vary in whether or not respondents must be citizens or be fluent in specific languages [21].

- An increasingly common form of housing seen in international studies is workers' quarters. Survey designers may want to explicitly state in the definition of the target population whether workers' quarters should be included or excluded.

3. **Identify and evaluate potential sampling frames. Select or create the sampling frame that best covers the target population given the country's survey budget.**

Rationale

An ideal sampling frame contains all of the elements of the target population. However, very few sampling frames exist that allow access to every element in the target population. The goal, then, is to choose a sampling frame or a set of sampling frames that allows access to the largest number of elements in the target population and contains the fewest number of ineligible elements, given the constraints of the survey budget.

Procedural steps

- Have each participating country identify a pre-existing list (or lists) of desired elements or clusters of elements of the target population to create a sampling frame. Examples include:
  - Official population registries.
  - Postal registries.
  - Electoral rolls.
  - Pre-existing sampling frames used by other surveys.
  - Telephone directories.
  - Other list(s) of addresses, phone numbers or names.
    - Although this chapter focuses heavily on the method and practice of in-person interviews, telephone interviews using Random-Digit-Dialing (RDD) frames [39] or other lists are also widely used. In cross-national surveys, the situation may occur where one country conducts interviews over the telephone while another country conducts face-to-face interviews. This difference in the mode of data collection, driven by the available sampling frames, might lead to differences in the results. (See Data Collection for more information about mode effects.)

- Create a sampling frame via area probability sampling methods if there are no appropriate pre-existing lists of elements of the target population; even if such lists do exist, it is wise to assess the cost and coverage errors associated with creating an area probability sampling frame. Many texts and documents provide detailed guidance regarding the development of area probability samples [23] [41]. Below, we outline a simple two-stage area probability sample of households, including the following steps used in many cross-cultural surveys. Additional information can be found in Appendices A and C:

- Create a list of primary sampling units (PSUs) based on geographic clusters. In the United States, for example, these clusters are typically census enumeration areas.
- Using a probability sampling method, select a sample of PSUs.
- Determine the appropriate method for listing the housing units (secondary sampling units (SSUs)) within selected PSUs.
- Send staff to list the housing units in selected PSUs, maintaining a uniform definition of what constitutes a "housing unit."
- Once the housing units in a PSU have been enumerated, select a random sample of housing units from the list.
- During data collection, ask the selected housing units within the PSUs to participate. Once the housing unit has agreed to participate, complete a list of all eligible members within the housing unit. (See Appendix B for more detailed instructions on enumerating eligible members of the housing unit.)
- Using a probability method, select one or more eligible members within the housing unit.
    - Train the interviewer or, where possible, program the computer to select an eligible respondent based on the selection method specified.
    - While some "quasi-probability" and "non-probability" or "quota" within-household selection methods can be used, be aware that such procedures produce a non-probability sample.
    - Some studies may want to survey the most knowledgeable adult, the one with primary child care responsibilities, or with some other specific characteristics, rather than randomly select from among the household members. Note that this is part of the definition of the target population and, thus, does not violate probability sampling.

- Evaluate how well each potential sampling frame covers the target population [9]. (For more information, refer to Appendix C.)
    - Examine the sampling frame(s) for a one-to-one mapping between the elements on the sampling frame and the target population. There are four potential problems:
        - Undercoverage (missing elements): elements in the target population do not appear on sampling frame.
        - Ineligible elements: elements on the sampling frame do not exist in the target population.
        - Duplication: several sampling frame elements match one target population element.
        - Clustering: one sampling frame element matches many target population elements.
    - Area frames generally have better coverage properties than pre-existing lists of addresses, names, or telephone numbers because area frames have fewer missing eligible elements, fewer

duplications, and fewer ineligible elements. (For more information on the creation of area probability frames, see Appendices A and C.)

- Consider combining multiple sampling frames which cover the same population to create a list of sampling units if the union of the different frames would cover the target population better than any one of the frames on its own [15]. When combining multiple lists to create a sampling frame, the following steps should be considered [12]:
  - First, determine for each element on the combined frame whether it is a member of Frame A only, Frame B only, or both Frame A and B. To calculate the joint probability of an element being selected from Frame A or Frame B, use the following formula:

$$P(\mathrm{A}+\mathrm{B}) = P(\mathrm{A}) + P(\mathrm{B}) - (P(\mathrm{A}) * P(\mathrm{B}))$$

    - If the membership of each element can be determined before sampling, duplicates can be removed from the sampling frame.
    - A variation on this is to use a rule that can be applied to just the sample, rather than to the entire frame. Frame A might be designated the controlling frame, in the sense that a unit that is in both frames is allowed to be sampled only from A. After the sample is selected, determine whether each unit from B is on the A frame, and retain the unit only if it is not on frame A. This method extends to more than two frames by assigning a priority order to the frames.
    - If the membership cannot be determined prior to sampling, then elements belonging to both frames can be weighted for unequal probabilities of selection after data collection (see Data Processing and Statistical Adjustment for best practices for weighting and nonresponse adjustments).

- Assess the cost of obtaining or creating each potential sampling frame.
  - In most circumstances, it is less expensive to purchase pre-existing lists than to create area probability frames.
  - While three stage area probability samples are more costly to develop than pre-existing lists, they facilitate cost-effective clustering for interviews.
  - If the pre-existing lists are not up-to-date, potential respondents may no longer live at the address on the list or may have changed phone numbers; tracking these individuals can be very expensive.
  - Pre-existing lists for household surveys often contain more ineligible elements than area probability frames, increasing survey costs.

- Update an already existing frame, if necessary. For example, World Health Survey (WHS) administrators have suggested that frames that are two years old or more require updating [41]. However, that is only a rough rule of thumb. In mobile societies, two years might be too long while, in rather static societies, even older frames might still be accurate.
  - If the frame is a pre-existing list, contact the provider of the list for the newest version and its quality documentation.
  - If the frame is an area probability sample and the target population has undergone extensive movement or substantial housing growth since the creation of the frame, then updating the PSUs and SSUs will be required. However, what is most important is the quality of the enumerative listing.
  - Select the sampling frame based on the undercoverage error vs. cost tradeoff.

Lessons learned

- Most countries do not have complete registers of the resident population and, therefore, construct area frames for sample selection. Some surveys in majority countries have found that it can be difficult to enumerate the rural, poor areas [2] [8] [18] and, consequently, surveys in these countries may under-represent poorer or more rural residents. (Not all survey methodologists agree with the opinions expressed by these authors regarding enumeration in rural, poor areas. Those who disagree argue that the poor enumerations are mainly due to low expectations and insufficient training and supervision.) If the statistic of interest is correlated with income and/or urbanicity, the sample estimate will be biased. For example, the Tibet Eye Care Assessment, a study on blindness and eye diseases in the Tibet Autonomous Region of China, used an area sampling frame [8]. One of the PSUs was the township of Nakchu, an area of high elevation that is primarily populated by nomadic herders. Because of the elevation and rough terrain, Nakchu proved difficult to enumerate accurately. As a result, the survey sample underrepresented the residents of the roughest terrain of Nakchu. This was potentially important, as ophthalmologists believe that Tibetans who live in the most inaccessible regions and the highest elevation have the highest prevalence of eye disease and visual impairment.

- Local residents can help produce maps for an area probability sample. When measuring the size of the rural population in Malawi, researchers used statistical methods to determine the sample size and selection of villages. Then they asked members of the selected communities to help draw maps, including exact village boundaries, key landmarks, and each individual household [1].

4. **Choose a selection procedure that will randomly select elements from the sampling frame and ensure that important subgroups in the population will be represented.**

Rationale

Sample selection is a crucial part of the survey lifecycle. Since we cannot survey every possible element from the target population, we must rely on probability theory to make inferences from the sample back to the target population.

Procedural steps

- Consider only selection methods that will provide a probability sample.
  - Statisticians have developed procedures for estimating sampling errors in probability samples which apply to any type of population.
  - Random sample selection protects the researcher against accusations that his or her bias, whether conscious or unconscious, affected the selection.
  - Creating a frame where each element has a known, nonzero probability of selection can, in some cases, be very costly in terms of both time and effort. To reduce costs, some survey organizations select nonprobability samples such as convenience samples (sampling units are selected at the convenience of the researcher, and no attempt is made ensure that the sample accurately represents the target population) or quota samples. Upon the completion of data collection with such a sample, the survey organization typically calculates population estimates, standard errors, and confidence intervals as though a probability sample had been selected. In using a nonprobability method as a proxy for a probability method, the survey organization makes the assumption that the nonprobability sample is unbiased. While not all nonprobability samples are biased, the risk of bias is extremely high and, most importantly, cannot be measured. A survey that uses a nonprobability sampling method cannot estimate the true error in the sample estimates [16].

- Identify the optimal sampling method available in each country. (Below are summaries of each selection method. See Appendix D for additional information about each selection method.)

- Consider Simple Random Sampling (SRS) without replacement. In SRS, each element on the frame has an equal probability of selection, and each combination of *n* elements has the same probability of being selected. Due to the benefits of stratification, this technique is seldom used in practice.

- Advantages of SRS:
  - The procedure is easy to understand and implement.
- Disadvantages of SRS:
  - The costs in attempting to interview a simple random sample of persons can be quite high.
  - SRS provides no assurance that important subpopulations will be included in the sample.

- Consider Systematic Sampling to reduce the operational effort needed to select the sample. In systematic sampling, every *k*th element on the sampling frame is selected after a random start.
  - Advantages of systematic sampling:
    - The operational time necessary to select the sample can be reduced substantially.
    - If the sampling frame is sorted into groups or ordered in some other way prior to selection, the systematic sampling method will select a proportionately allocated sample (see description below of stratified sampling). This is often referred to as "implicitly stratified sampling."
  - Disadvantages of systematic sampling:
    - If the key selection variables on the sampling frame are sorted in a periodic pattern (e.g., 2, 4, 6, 2, 4, 6…) and the selection interval coincides with periodic pattern, the systematic sampling method will not perform well [18]. If periodicity is a problem, several systematic samples can be selected and concatenated to form the total survey sample.
    - If the list is sorted in a specific order before selection, the repeated sampling variance of estimates cannot be computed exactly.

- Consider Stratified Sampling (see Appendix D for a detailed description). Stratified sampling uses auxiliary information on the sampling frame to ensure that specified subgroups are represented in the sample and to improve survey precision. Virtually all practical sampling uses some form of stratification.
  - Advantages of stratified sampling:
    - Depending on the allocation of elements to the strata, the method can produce gains in precision (i.e., decrease in sampling variance) for the same efforts by making certain that essential subpopulations are included in the sample.
  - Disadvantages of stratified sampling:
    - Selection of stratification variables that are related to the statistic(s) of interest can sometimes be difficult.

    - No gains in precision will be seen if the stratification variables are not correlated with the statistic(s) of interest. In very rare cases, the precision may even decrease.

- Consider Cluster Sampling (see Appendix D for a detailed description). With cluster sampling, clusters of frame elements are selected jointly rather than selecting individual elements one at a time. The only population elements listed are those within the selected clusters.
  - Advantages of cluster sampling:
    - When survey populations are spread over a wide geographic area and interviews are to be done face-to-face, it can be very costly to create an element frame and visit *n* elements randomly selected over the entire area.
    - A full frame of all elements in the entire population is not required.
  - Disadvantages of cluster sampling:
    - Estimates are not as precise as with SRS, necessitating a larger sample size in order to get the same level of precision (See Appendix E for more information about effective sample size).

- Consider Two-Phase (or Double) Sampling (see Appendix D for a further description). The concept of two-phase sampling is to sample elements, measure one or more variables on these 1st-phase elements, and use that information to select a 2nd-phase subsample.
  - A common application is to collect 1st-phase data that is used to stratify elements for the 2nd-phase subsample.
  - Survey samplers use two-phase sampling to help reduce nonresponse, with the stratifying variable from phase one being whether the person responded to the initial survey request. For example, samplers might select a subsample of nonrespondents and try to entice the nonrespondents to participate by offering incentives.

- Consider Replicated (or Interpenetrated) Sampling. Replicated sampling is a method in which "the total sample is made up of a set of replicate subsamples, each of the identical sample design [18]."
  - Advantages of replicated sampling:
    - It allows the study of variable nonsampling errors, such as interviewer variance.
    - It allows for simple and general sampling variance estimation (see Data Processing and Statistical Adjustment for further explanation, especially Balanced Repeated Replication and Jackknife Repeated Replication).
  - Disadvantages of replicated sampling:

  - In face-to-face surveys, random assignment of interviewers to areas, rather than assignment to geographically proximal areas of the country, can lead to very large increases in survey costs.
  - There is a loss in the precision of sampling variance estimators; a small number of replicates leads to a decrease in the number of degrees of freedom when calculating confidence intervals.

- Consider using a combination of techniques such as a stratified multistage cluster design.
  - Most surveys in majority countries are based on stratified multistage cluster designs [43]. The combination of these techniques reduces data collection costs by clustering while striving to increase or maintain precision through stratification.

Lessons learned

- Probability sampling at every stage generally requires more labor and funding than other methods. Therefore, some cross-cultural studies have used probability sampling in the first stage of selection, then allowed quota sampling or substitution to occur at later stages [4] [16]. However, a survey that uses a nonprobability sampling method at any stage of selection cannot estimate the true error in the sample estimates [16]. Therefore, the coordinating center should make every effort to promote the use of a full probability sample and remove any obstacles that would prevent participating countries from using probability methods at each stage of selection. For the first few waves of data collection, the International Social Survey Programme (ISSP) allowed countries to use nonprobability methods at the household level. After recognizing the problem this caused in variance estimation, the ISSP has required countries to use full probability samples since 2000 [14].

- Existing cross-cultural surveys have employed various strategies for selecting a probability sample.
  - Round 4 of the Afrobarometer Survey uses a clustered, stratified, multi-stage, area probability sample. The sampling design has four stages: (1) stratify and randomly select primary sampling units, (2) randomly select sampling start-points, (3) randomly choose households, and (4) randomly select individual respondents within households [44].
  - Sample designs vary across participating countries in the Asian Barometer, but all are sampled with probability proportional to size [45].
  - Samples for Round 5 of the European Social Survey (ESS) must use random probability sampling at every stage. Samples are designed by a sampling expert or panel and may include clustering

and stratifying. Quota sampling and substitutions are not allowed although subgroups may be over-sampled. Sample designs and frames must be documented in full and be pre-approved by a sampling expert or panel. The target minimum response rate is 70% and the maximum non-contact rate is 3% [47].

- Sampling frames and designs for the Living Standard Measurement Study Survey (LSMS) vary across participating countries but generally consist of two stages. In the first stage, the sample frame is developed from census files and Primary Sampling Units are randomly selected with probability proportionate to size; in the second stage, households (usually 16) are randomly selected from each of the designated Primary Sampling Units. Clustering and stratifying are permitted, but all sampling procedures must be documented and made available to data analysts [27].
- Survey of Health, Aging and Retirement in Europe (SHARE) sampling designs vary by country but all are required to be probability samples. Three sampling designs may be used: (1) stratified simple random sampling from national population registers, (2) multi-stage sampling using regional or local population registers, or (3) single or multi-stage sampling using telephone directories followed by screening in the field [49].
- Sampling frames for the World Mental Health Survey vary across participating countries, but generally consist of three types of sampling frames: (1) individual contact information databases such as national population registries, voter registration lists, or household telephone directories, (2) multistage area probability sample frames, or; (3) hybrid multistage frames that combine area probability methods and a individual contact database in the final stages. Sampling designs vary across participating countries, including stratification and clustering, but probability sampling is required at all stages. The target minimum response rate is 65% [21].
- Probability sampling is strongly recommended, but not required, in the World Value Survey; any deviations from probability sampling are to be reported in the Methodology Questionnaire report [42] [51].

5. **Determine the sample size necessary to meet the desired level of precision for the statistics of interest at population or subgroup levels for the different potential sample selection procedures.**

Rationale

After choosing a sample design, and before selecting the sample from the sampling frame, the sample size must be determined. The sample size takes into account the desired level of precision for the statistic(s) of interest, estimates of the statistic of interest from previous surveys, the design effect, and estimated outcome rates of the survey. (See [29] for a detailed treatment of the approach used in the European Social Survey. For a more extensive example of sample size calculation, see Appendix E.)

Procedural steps

- Specify the desired level of precision, both overall and within key subgroups. Practical experience has determined that often it is easiest for sponsors to conceptualize desired levels of precision in terms of 95% confidence intervals.

- Convert these 95% confidence intervals into a sampling variance of the mean or proportion.

- Obtain an estimate of S^2 (population element variance).
  - If the statistic of interest is not a proportion, find an estimate of S^2 from a previous survey on the same target population or from a small pilot test.
  - If the statistic of interest is a proportion, the sampler can use the expected value of the proportion (p), even if it is a guess, to estimate S^2 by using the formula $s^2 = p(1-p)$.

- Estimate the required number of completed interviews for a SRS by dividing the estimate of S^2 by the sampling variance of the mean. (See [7] for more on sample size computation for SRS.)

- Multiply the number of completed interviews by the design effect to account for a non-SRS design.

- Calculate the necessary sample size by dividing the number of completed interviews by the expected response rate, eligibility rate, and coverage rate.
  - The sampler can estimate these three rates by looking at the rates obtained in previous surveys with the same or similar survey population and survey design.

Lessons learned

- Prior to the first implementation of the European Social Survey (ESS), many of the participating survey organizations had never encountered the concepts of sample size determination and calculating design effects [29]. Therefore, the ESS expert sampling panel spent considerable time explaining these. In return, the organizations that were new to these methods were very enthusiastic to learn about them, and eager to meet the standards of the coordinating center. In fact, after completing Round 1 of the study, many nations commented that designing the sample was one of the most educational aspects of the entire survey process, and had significantly improved the survey methods within their country.

- Sample size sometimes varies among countries participating in cross-cultural surveys. In Round 4 of the Afrobarometer Survey, sample size ranges from a minimum of 1,200 respondents to 2,400 or more in extremely heterogeneous areas [44]; sample size ranges from 800 to 3,200 respondents in the Asian Barometer study [45]; Round 5 of the European Social Survey (ESS) requires a minimum of 800 respondents for participating countries that have a population of less than two million, 1,500 from larger countries [47]; the International Social Survey Programme (ISSP) requires a minimum of 1,000 respondents, with a goal of 1,400 respondents [48]; sample size ranges from 1,600 to 5000 households in the Living Standard Measurement Study Survey (LSMS) [27]; the Survey of Health, Ageing and Retirement in Europe (SHARE) requires 1,500 respondents from each participating country [49]; samples in the World Mental Health Survey range from 2,357 (Romania) to 12,992 (New Zealand) [21]; and the World Value Survey requires a minimum of 1,000 respondents [42] [51].

6. Institute and follow appropriate quality control procedures at each step of the sample design process.

Rationale

Development and implementation of quality control procedures for the sample design are necessary to ensure the highest level of coverage possible and to maintain a probability sample that meets the desired level of precision for key survey statistics. If a failure to meet those standards is detected, protocols should be in place to remedy the failure. In addition, monitoring of procedures related to the sample design of the study should inform efforts to improve the quality and cost-effectiveness of the study over time.

Procedural steps

- Define the target population for the study across all participating countries/cultures as well as the target population within each country/culture. If the study design does not change over time, strive to keep each target population, both overall and within participating country, consistent over time.

- Prior to selecting sample elements or sampling units, provide the data collection staff with a list of all of the variables on the sampling frame and ask which variables they would like and the format in which they would like these variables delivered once sampling is complete for data collection purposes.
  - After sample selection, check that each selected sampling unit or element contains this information and is in the specified format.

- If possible, use a responsive survey design [11] [13] to help achieve an optimal sampling design (see Survey Quality for more information about responsive survey designs). A responsive survey design uses prespecified paradata (quantitative indicators of the data collection process such as "contact attempts" or "interviewer success rate") for intervention during data collection.
  - Advantages of Responsive Survey Designs are the prespecification of interventions instead of ad hoc decisions and the possibility to target efforts on hard to interview groups.
  - A disadvantage is that the survey designers walk a thin line between full probability and quota if they deviate from carefully predefined paradata-driven interventions.

- After each stage of selection, generate frequency tables for key variables from the frame of sampling units to check for the following:
  - Overall sample size and within stratum sample size.
  - Distribution of the sample units by other specific groups such as census enumeration areas.
  - Extreme values.
  - Nonsensical values.
  - Missing data.

- Create a unique, sample identification code for each selected sampling unit. This code will allow identifying information to be easily removed after completing data collection.

○ Whether the participating country or the coordinating center is selecting the sample, assign a second sampling statistician within that organization to check the sample design methodology and the statistical software syntax of the survey's primary sampling statistician.

- Save all data files and computer syntax from the preferred statistical software package needed during sample design process in safe and well-labeled folders for future reference and use.

Lessons learned

- The construction and maintenance of sampling frames constitute an expensive and time-consuming exercise. If a participating country determines that no sampling frame meeting the specified coverage level of the target population exists, they can create a frame from such sources as census data collected by national statistics offices. However, one should be aware that official statistics differ greatly in accuracy from country to country.

- As discussed in the Lessons Learned section of Guideline 2, the decision to stray from full probability sampling reflects the conflict between standardization and flexibility in cross-cultural surveys. However, it bears repeating that without probability sampling, one cannot make justifiable inferences about the target population from the sample estimates.

7. Document each step of the sample selection procedure.

Rationale

Over the course of many years, various researchers will analyze the same survey data set. In order to provide these different users with a clear sense of how and why the data were collected, it is critical that all properties of the data set be documented. In terms of the sample design and selection, the best time to document is generally shortly after sample selection, when the information regarding sample selection is fresh in one's mind.

Procedural steps

- Have participating countries document the sample selection procedure while selection is occurring or shortly thereafter. Ideally, set a deadline that specifies the number of days after sample selection by which each participating country must send sampling selection documentation to the host survey organization. Be sure to allow for appropriate time to review and revise documentation when setting the deadline. (See Tenders, Bids and Contracts.)

- Include the following:
  - A clear definition of the survey population, as well as the differences between the target population and survey population.
  - The sampling frame:
    - Both the sampling frame used and the date the frame was last updated, if the frame is a registry or list.
    - A description of the development of the sampling frame and the frame elements.

- A description of the how well the sampling frame is thought to cover to target population and the potential for coverage error.
  - The data file of selected elements:
    - A descriptive and distinct variable name and label.
    - Unique variables that contain the selection probabilities at each stage of selection as well as the overall selection probabilities. If a participating country used a nonprobability method in at least one stage of selection and therefore the selection probabilities are unknown, ensure that this is clearly documented.
    - A clear description of all variables in the selected element data file, with all variable names, an accompanying description and a codebook, which provides question-level metadata that are matched to variables in a dataset.
    - The statistical software syntax used for checking the dataset of selected sampling units or elements.

- For each sample, indicate how many stages were involved in selecting the sample (include the final stage in which the respondent was selected within the household, if applicable), and a description of each stage, including how many sampling units were selected at each stage.
  - Examples of different stages include:
    - State/province.
    - County or group of counties.
    - City/town, community, municipality.
    - Census/election district.
    - Area segment/group of neighborhood blocks.
    - Housing unit/physical address (not necessarily the postal address).
    - Postal delivery point/address.
    - Block of telephone numbers (e.g., by regional prefix).
    - Telephone number.
    - Household.
    - Person selected from a household listing.
    - Named person selected from a list, registry or other source that was *not* a household listing.
  - Examples of how sampling units might be selected:
    - All selected with equal probability.
    - All selected with probability proportional to size; specify the measure of size used. (See Appendix E for more on probability proportional to size sampling methods.)
    - Some units selected with certainty, others selected with probability proportional to size; describe the criteria used for certainty selection.
    - Census/enumeration (all units selected with certainty).

- Units selected using a nonprobability method (e.g., convenience sample, quota sample).
- At each stage of selection, describe the stratification variables and reasons for choosing these variables. Some examples of commonly used stratification variables are:
  - Age.
  - Region of the country.
  - State/province.
  - County.
  - City/town, community, municipality.
  - Postal code.
  - Metropolitan status/urbanicity.
  - Size of sampling unit (e.g., population of city).
  - Race/ethnicity.
  - National origin (e.g., Mexican, Nigerian).
- At each stage of selection, explain the allocation method used and the sample size for each stratum at each stage of selection. (See Appendix E for more on allocation methods in stratified sampling.)

- If systematic sampling was used at any stage of selection, indicate whether the frame was sorted by any variables prior to systematic selection in order to achieve implicit stratification. If this is the case, describe the variable(s).

- Describe the time dimension of the design (i.e., one-time cross-sectional, fixed panel, rotating panel design).
  - If a panel study:
    - State how many previous waves or rounds of data collection there have been for this panel study.
    - Describe the initial sample design for the panel study and any subsequent modifications to the design that are important in documenting this study.
  - If a rotating panel design:
    - Fully describe the rotating panel design for the study (e.g., fresh cross-section is drawn each month and respondents are interviewed once that month, and then reinterviewed once six months later).
    - State the anticipated precision of the estimates.
    - Explain any problems encountered during the sampling process and any deviations from the sampling plan during implementation.
  - Additional sampling documentation:
    - Report any (additional) subsampling of eligible respondents, carried out in order to control the number of interviews completed by respondents with particular characteristics (e.g., one in two eligible males was interviewed, one in four eligible

persons with no previous history of depression was interviewed (describe protocol)).

- Describe any use of replicates (see Data Processing and Statistical Adjustment).
- Explain if releases (nonrandom subsets of total sample) were used or the entire sample was released to data collection staff at the start of the study.
- Recount in detail any substitution or replacement of sample during data collection.

Lessons learned

- As the procedural steps outlined above show, selecting a sample can involve many detailed steps that may be hard to recall after the fact. For example, the coordinating center for the World Mental Health Survey began gathering sampling documentation for weighting and other purposes after many of the participating countries had finished data collection. They found that some countries had a difficult time recalling all the necessary details, such as the sample size for each stratum at each stage of selection. It is wise to document sampling procedures in detail shortly after sample selection (see Data Processing and Statistical Adjustment for further explanation of weighting practice).

Appendix A

Additional information on creating area probability sampling frames

Creating and selecting primary sampling units (PSUs)

- Create PSUs. PSU's are geographic clusters. In the United States, they are often census enumeration areas, postal codes, or election districts. The size of the geographic clusters should be large enough to contain a population that is heterogeneous with respect to the survey variables of interest, but small enough to realize the travel-related cost efficiencies of clustered sample observations. Good PSUs generally have the following characteristics:
  - They possess clearly identifiable boundaries that are stable over a certain time. (Note that all administrative boundaries such as census enumeration areas, election districts, etc., are regularly updated and changed.)
  - They cover the target population completely.
  - They have measures of size for sampling purposes.
  - They contain auxiliary data for stratification purposes (see Guideline 3).
  - They are large in number.

Defining and enumerating secondary sampling units (SSUs)

- Decide on a comprehensive definition of a housing unit (HU).
  - What defines a HU and who should be counted as a household member can vary greatly across countries. For comparative surveys, often only a general definition is feasible (e.g., all persons living in private households born before xx/xx/xx in country y). Be aware that the size of a typical "private household" also varies among countries. Hoffmeyer-Zlotnik and Warner provide many household definitions used in the European Union [17].
  - A commonly used definition in the United States is "a physical structure intended as a dwelling that has its own entrance separate from other units in the structure and an area where meals may be prepared and served [9]."
  - In 1998, the United Nations defined a HU as "a separate and independent place of abode intended for habitation by a single household, or one not intended for habitation but occupied as living quarters by a household at the time of the census. Thus it may be an occupied or vacant dwelling, an occupied mobile home or improvised HU, or any other place occupied as living quarters by a household at the time of the census. This category thus includes housing of various levels of permanency and acceptability" [40].

- Determine the appropriate method for listing the HUs (SSUs) within selected PSUs.
  - One option is to use a preexisting list of HUs.
    - Some cross-cultural surveys have used satellite technology to help identify and list households, settlements, and habitations, especially those in hard to find areas such as mountainous, riverine, and creek regions [32].
  - Another option is to send staff to list the HUs in selected PSUs.
    - Create maps to help staff efficiently travel to and correctly list all of the HUs within the selected PSUs. (See the section below on maps for creating two stage area probability frames of HUs.)
    - Use standardized protocol to consistently enumerate the HUs in selected PSUs in the field.
    - If a preexisting list of HUs for the specified PSU is available but the list is believed to be incomplete or if the coverage properties of the list are unknown, the participating country can send staff to the PSU with the pre-existing list and instructions to update and revise the list so that all HUs are included.
    - If no pre-existing list is available or the participating country knows from previous experience that the available list greatly undercovers the HUs in the PSU, have staff enumerate all the HUs in the PSU without the aid of a list.
    - If some selected PSUs have lists of HUs that, at least marginally, cover all its HUs and other PSUs do not, a combination of these listing methods can be used.

Creating maps to help staff locate PSUs and enumerating SSUs

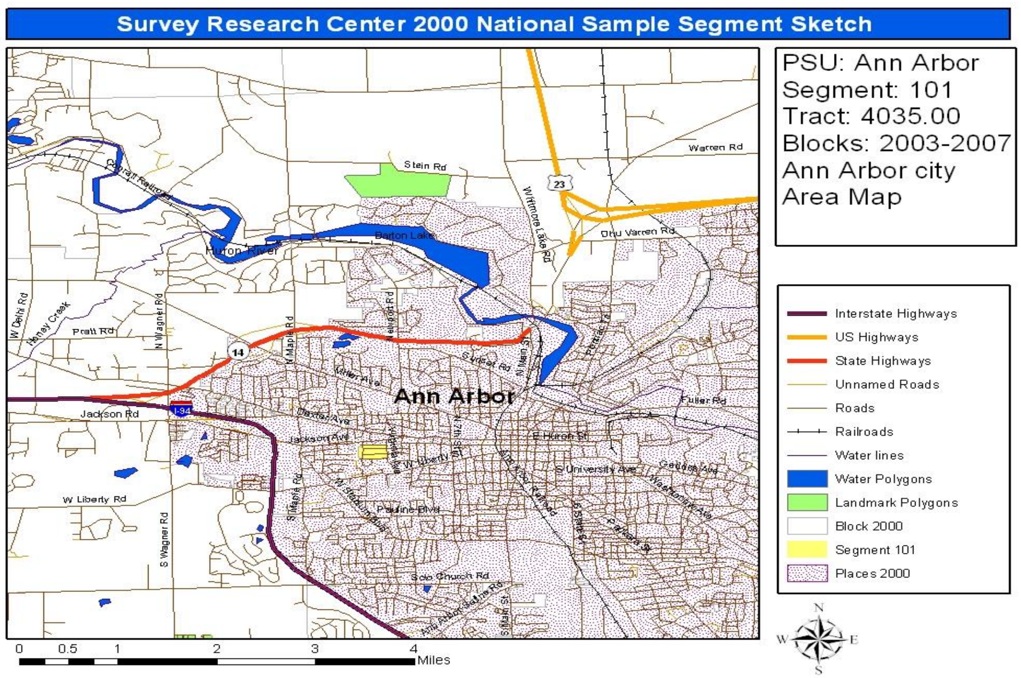
- Most of the surveys conducted in majority countries are based on multistage, stratified area probability sample designs. The example here is for a two-stage area probability design of HUs where the PSUs are groups of linked United States Census blocks and the SSUs are HUs (both occupied and unoccupied) within the selected blocks. Maps can be created on paper by hand or electronically with a mapping program like ArcGIS [30] that uses geographic data. Likewise, maps may be distributed on paper or electronically.

Area Maps

- The purpose of the area map is to show a geographic area large enough to provide context for locating the selected PSUs. Useful area maps typically contain the following features:
  - Map Layout: Create area maps so that the top of the map indicates north and the top right corner of the map page displays the name of survey areas, and their associated area numbers.

- Map Legend: Located under the area information, the legend identifies roads, streets, and highways. Water boundaries for creeks, streams, rivers and lakes can be coded blue. Railroads can be indicated with a cross-hatched line.
- Distance Scale: At the bottom of the map, a scale indicates the range of miles/kilometers the map encompasses.

Example of an area map [35]



PSU Maps

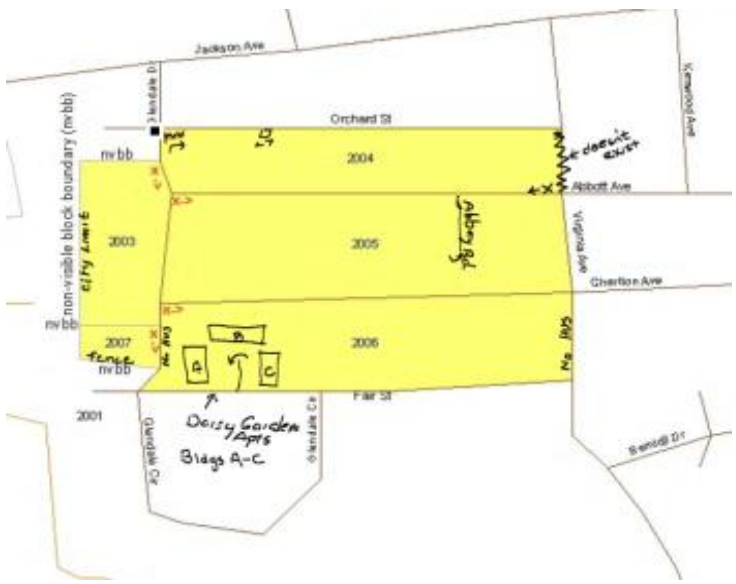
- The purpose of a PSU map is to update or correct street names, note the line of travel used when listing, and draw landmarks or physical boundaries that will help future interviewers find all the listed HUs in the PSU. Below are detailed instructions for creating PSU Maps:

- Starting X and Directional Arrows: Draw a starting X and directional arrows to assist with the listing assignment. Make an effort to determine a logical starting place for listing each block, like the Northeast corner.

- Non-Visible Block Boundaries (NVBB): If listers are having difficulty locating an NVBB, check with town or city officials for the exact location of the line and then "pin" it down on the map.

- Once listers have visited the PSU in person, best practice suggests that they update the map to accurately reflect the defined geographical area including:
  - Obtaining information about new streets or housing construction. A visit to the city or county planning office, or the engineering or highway department in the area can usually provide the information to accurately record current conditions in the segment area.
  - Recording "no household units" along any block face that is clearly devoid of HUs, such as those with parks, vacant fields, parking lots, woods, farm land, or only commercial or industrial structures.
  - Recording street names missing from the map(s), drawing in streets, alleys or cul-de-sacs not shown on the map, and correcting misspelled or incorrect street names. Verify that street names are complete.

Example of PSU map [35]

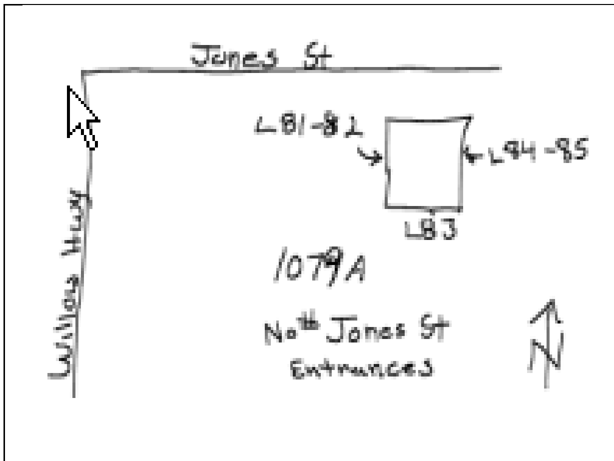


Sketch Maps

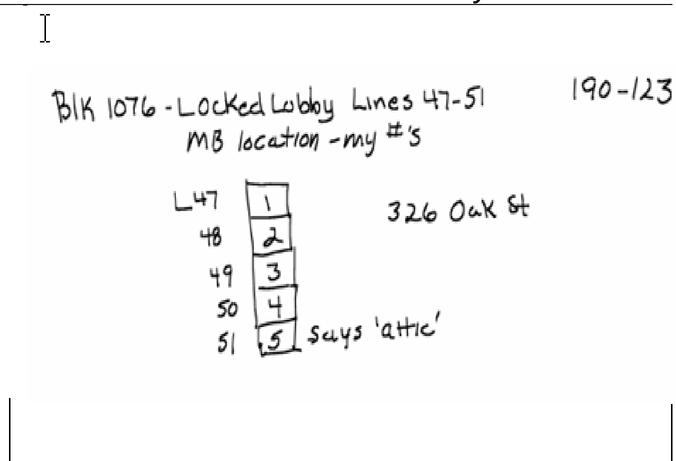
- The purpose of a sketch map is to allow listers to supplement the PSU map with their own hand-drawn map when the area and PSU maps provided seem inadequate.
- A few examples of sketch maps are provided below but the list is not all-inclusive [35]. Sketch Map 2 is an example of sketch map used in rural area in China in the China Health and Retirement Longitudinal Study (CHRLS). For more information on the methods for listing in areas without street addresses, rural, or unmapped areas, refer to [36] and [37].

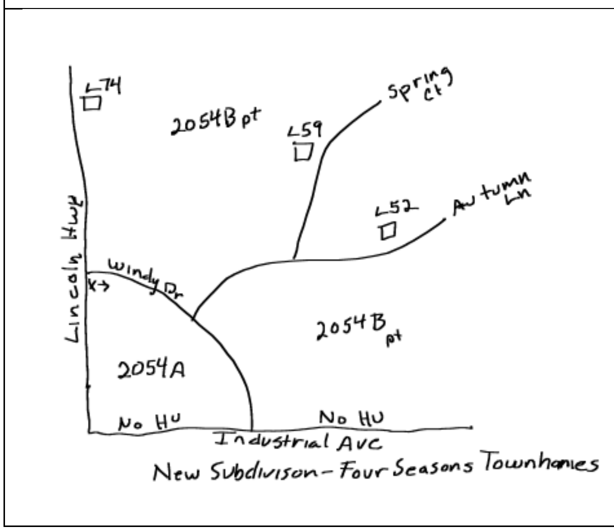
Sketch Map 1: An example of a sketch map used in a survey in the United States [35]

Sketch of exterior door locations

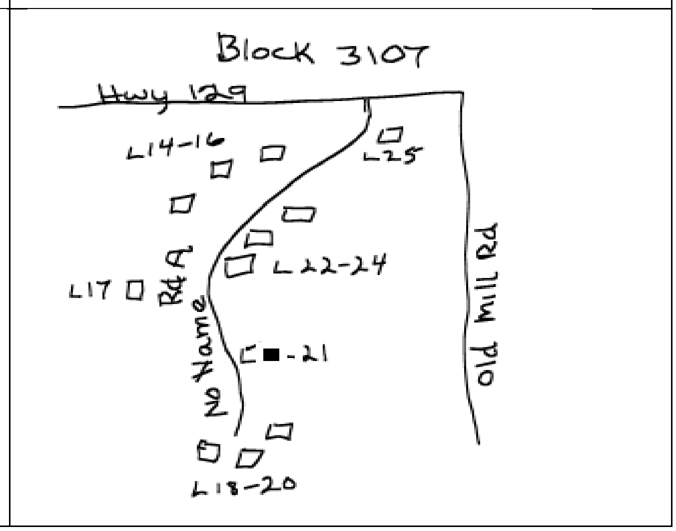


Sketch of mailboxes in locked lobby



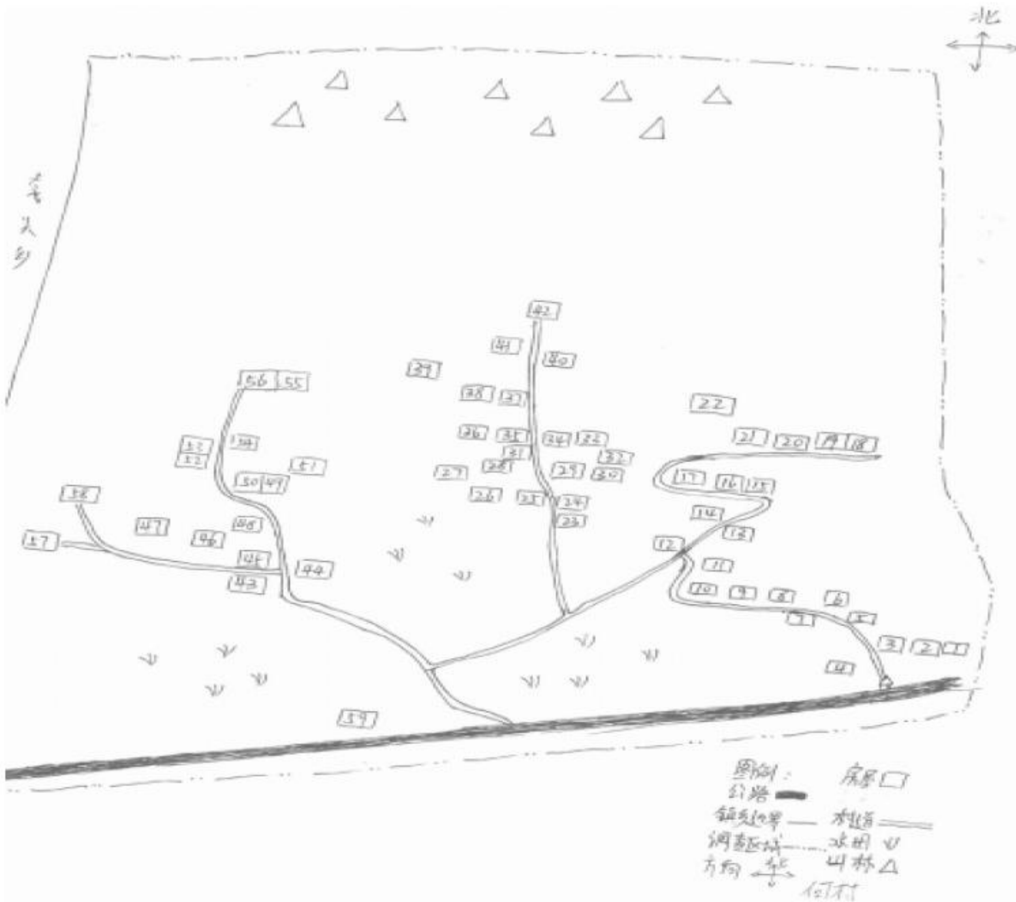


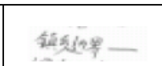
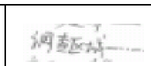
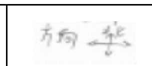
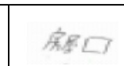
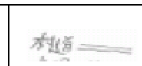
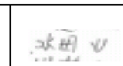
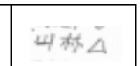
Sketch of street configuration for a new subdivision



Sketch of "No Name" Road

Sketch Map 2: An example of a sketch map used in the China Health and Retirement Longitudinal Study (CHRLS) [5]



							
Road	Village Boundary	Survey Area	Direction	House Unit	Village Street	Paddy Field	Forest Area

<u>Tasks for listing staff to complete prior to enumeration of HUs</u>

- Train listers to complete the following tasks prior to beginning the listing procedure:
  - Contact local authorities. A survey organization can provide listers a letter and a form to deliver to the local police station or some other local authority, alerting them to the survey's presence in their area. The letter to the local authorities might define the purpose for listing efforts and also give staff a chance to gather information about the local situation.
  - Scout the selected areas. Most experienced listers make a complete circuit of selected areas once before beginning listing to get a "feel" for the area. The purpose is to help find the assigned

areas and to confirm or correct boundaries if maps obtained are hard to read.

- It is helpful if listing staff estimate the number of HUs and look for indicators that may explain the discrepancy if it appears that there are twice as many HUs or fewer than half as many HUs as expected by the census count (e.g., new apartment complex or subdivision, HUs which have been demolished in the recent past, or older homes which have been converted to other uses).

Recording the listed HUs

- Listing is an exact record of all HUs, both occupied and unoccupied, that are located in predefined census geographical area boundaries.

- Elements of Listing Format for United States Hus include:

- Block: borough, planned residential area or village number.
- Non mailable (NM) indicator which is used to indentify addresses that cannot receive postal mail because the address is not complete or not unique.
- Line Number (Line_No): the first HU recorded in every listing begins at Line Number 1 with the subsequent HUs encountered being numbered consecutively 2, 3, 4 and so on through all HUs found in the PSU.
- Street Number (Street_No): the street address number should be complete.
- Street Name: check the spelling of street names on sign posts against the street names given on the PSU maps.
- Apartment Number (Apt/Lot): this field should be used for apartment or trailer lot numbers only.

Example of listing format

Block	NM	Line_No	Street_No	Street Name	Apt/Lot	Additional Information
3003	X	1	NO#	FIRST AVE		
3003		2	654	FIRST AVE		
3003	X	3	1233	WILSON ST		
3003	X	4	1233	WILSON ST		Same Street_No as Line_No 3
3003		5	1241	WILSON ST		
3004	X	6	NO#	WILLOW HWY		

Additional protocols to help create consistent listings

- Make HUs listings consistent across all selected areas. The suggested listing protocols also include:
  - Begin by listing the lowest numbered block first. Work systematically around the PSU, listing each block in numerical order from lowest to highest block number.
  - Start listing HUs for each block beginning at the red starting “X” and following the directional arrow indicated on the Block Map.
  - Look “over your right shoulder” and record each HU address as it is approached. In other words all listed HUs should be on your right.
  - Walk around the block in a clockwise direction making a complete circuit until reaching the original starting point.
  - List only HUs inside the selected (shaded) PSU boundary.
  - List empty, boarded up, burned or abandoned HUs unless the HU is posted for demolition.
  - List on foot whenever possible when you are working in urban and suburban areas.

- Create general rules to deal with the following situations:
  - Abandoned, boarded up, burnt out, and vacant HUs.
  - Apartment complexes.
  - Locked buildings and gated communities.
  - New construction.
  - Under construction or unfinished construction.

- Check the completed listing.
  - Review the listed addresses against the block map for each block in the PSU. Beginning at the starting “X” and proceeding clockwise around each block, confirm that there are HUs listed for every street in the block or that HUs without a street number have been noted along the proper block face on the PSU map.
  - Confirm that there is only one HU per listed line and that listing lines are used only for HUs. Commercial or public buildings such as churches, schools, or businesses should be recorded only in the PSU observations or noted on the map(s).

- Make certain that all HUs without a street number are uniquely described in the additional information column and that their locations are noted on the map by line number.

    - Review the PSU observations and make sure they are complete. Confirm that information about locked building, seasonal accessibility, and safety issues are noted in detail.

- For example, street numbers may be transposed (3547 Main St should be 3574); a HU that was originally listed as not having an street number now has an street number posted over the front door; street names may be spelled incorrectly; directional indicators (N, SW, etc.) may be missing from the street name; or the street suffix (Ave, Hwy, Ln, St) may be missing.

Appendix B

Administration of the within housing unit (HU) listing of eligible persons

- Identify the eligibility of the selected HUs by listing the eligible persons within each selected HU (list of household members).

- Choose a HU residency rule to identify eligible respondents within each HU. Similar to defining a target population, once the rule is defined, it should be consistent across all participating countries. Choose between:
  - *De facto* residence rule – persons who slept in the HU the previous night.
    - Advantage: Easy to remember.
  - *De jure* residence rule – persons who "usually" sleep in HU.
    - Advantage: A better representation of the typical residents of a HU.
  - Design a household enumeration table based on study-specific residence rules and goals.
  - There are at least two sources of within-household undercoverage [31] [38]:
    - Motivated misreporting (deliberate concealment): household reporters deliberately conceal members for a multitude of reasons, including fear that they or another member may be evicted or deported.
    - Poor fit between living situation and definition: membership is complex and shows that household members may have confusion or disagree about who is a member.

Example 1 of a within household listing table

	HOUSEHOLD LISTING					RESPONDENT SELECTION		
	11 a. Household Member's First Name	11 b. HH Member's Relationship to Informant	11 c. Sex	11 d. Year of Birth	11 e. Language Spoken	11 f. Eligible	11 g. Person Number	11 h. Selected R
MALES			M					
			M					
			M					
			M					
			M					
			M					
			M					
FEMALES			F					
			F					
			F					
			F					
			F					
			F					
			F					

Instructions for using the household listing table

Column 11a (Household Member's First Name): List all members of the household, beginning with the informant. Note that males are listed in the upper portion of the table and females in the lower portion.

Column 11b (Household Member's Relationship to Informant): Record each household member's relationship to the informant (e.g., husband or wife, son or daughter, mother or father, brother or sister, friend, etc.).

Column 11d (Age): Record each household member's age.

Column 11e (Language Spoken): This column may or may not be included, depending upon the study requirements.

Column 11f (Eligible): Place a check mark in this column if, based upon the information in columns 11a-11e, the household member meets the eligibility criteria for the study.

Column 11g (Person Number): Assign a sequential number to each eligible household member.

Column 11h (Selected R): Count the number of eligible persons in the household. Find that number in the Kish table in the "If the Number of Eligible Persons is:" column. The selected respondent will be the household member with the "Person Number" corresponding to the "Interview the Person Numbered:" column in the Kish table (For more information about Kish tables, see Data Collection).

- Example 2 below is the 2010 Chinese Family Panel Study enumeration table. This study found the main challenge of listing to be situations where urban and rural villages were adjacent to one another. These situations contained complicated building structures and mixed populations (part-time and nonresident population). Therefore, the table specifically documents when more than one HU was located within a single dwelling, the reason the registered person had moved out, the time when the registered person moved out, and where the registered person had moved.

Example 2 of household enumeration table [6]

Tile: ____________
City: ____________Village Number: ____________ Lister: ____________ : Supervisor: ____________

Order	Name of HU Head	Registered Address	Registered order of Resident (Person)	Actual order of Resident (Person)	The person has lived here LESS than six months (please check)	If the person has moved out, please fill out this column. Reason (choose one) 1. Marriage. 2. Living with other relatives. 3. Moving to another place. 4. Having business at other place. 5. Not actually living but having registered record here. 6. Others.	If the person has moved out, please fill out this column. Where this person is living? 1. The same city 2. The same county 3. the same province 4. Out f this province	Eligibility (Check if the person is eligible)	HU Number (Assigned by supervisor)	Additional note
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)	(10)	(11)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
NO. of HUs in this address										

Appendix C

Additional Coverage Information

- Excluding ineligible elements.
  - Before the sample selection, these elements can be easily removed from the sampling frame. After sample selection, a question is needed, such as, "Is this structure a housing unit (HU)?" or, "Is this telephone number a household number?" Exclude all ineligible elements found after selection from the survey dataset.

- Removing duplicate elements from the sampling frame or accounting for them by weighting.
  - Duplicate elements can be easily removed from the sampling frame prior to sample selection. After selection, a question is needed, such as, "Does this HU have more than one household?" or, "Is there more than one telephone number serving this household?" Weight duplicated elements by the inverse of their chance of selection (See Data Processing and Statistical Adjustment).

- Identifying clustered elements prior to selection.
  - It is nearly impossible to identify clustered elements before selection. After sample selection, a question is needed, such as, "How many adults live in this household?" Weight all clustered elements by the inverse of their probability of selection (see Data Processing and Statistical Adjustment).

- Ways to reduce undercoverage [10].
  - Redefine the target population.
    - If all participating countries are having the same difficulty in capturing a subgroup of the target population (e.g., persons living in military bases and prisons), redefine the target population so those groups are excluded across all countries.

- Half-Open Interval (a heavily used technique in area probability samples).
  - Procedure most often used to help correct for missed HUs in the original listing of an area probability frame. The following are the steps to implement this technique:
    - Construct the frame in a prespecified order.
    - Define a system that links the ordered elements on the frame in either a forward (top-down) or backward (bottom-up) direction. Make the list circular, so that the first and last units on the list are linked.

- Sample from the frame.
- For each sampled element, check for missing elements between the selected element and the next linked element on the sampling frame. For example, when using an area probability frame, the survey organization will instruct interviewers to look closely for any missed HUs between the selected HUs and the next one on the list, before attempting to interview the selected HU.
- If missing elements are found, include these elements as part of the sample selection.
- If a large number of missing elements are found (e.g., a new apartment complex in between a sampled HU and the next HU on the frame), select a subsample of the missing elements.
- The major weakness of the Half-Open Interval is that it increases the uncertainty of the sample size.

- Using Network or Multiplicity Sampling techniques.
  - Network Sampling (a method to connect missing elements to elements on the frame, using a network rule)
  - This method can help cover hard-to-reach populations such as the homeless, but is rarely used in practice.
  - Steps:
    - Choose the most efficient sampling frame.
    - Draw a sample.
    - Devise a well-defined network (e.g., families defined by blood relations and adoption).
    - For each sampled element, gather a list of all eligible members of its network and select all or a subsample of the network members.

- Two methods to deal with clustering:
  - Select all eligible sampling units within the cluster.
  - Select a subsample of sampling units from within each cluster. For example, surveys often need to randomly select one or multiple people from within a household. (See Data Collection for more information on within-household respondent selection probability methods.)
    - Within-household respondent selection probability methods: A full listing of all eligible household members is created and one member selected at random. Some surveys may choose the most knowledgeable adult.
    - Kish grid with age and gender controls: Interviewers ask for the ages of all males who live in the household and then for the ages of all females who live in the household. Interviewers

consult selection tables developed by Kish [22] that select an adult member of the household depending on the number of adult males and females.

- Stratified probability-based subsampling with unequal probabilities of selection (e.g., white males over the age of 35 might have a probability of selection of 0.17 while African-American females under the age of 17 might have a probability of selection of 0.84): Survey data are weighted to account for unequal probabilities of selection.

- Other quasi-probability or non-probability methods, such as last birthday, youngest male, or convenience samples, should be avoided to maintain the ability to make accurate inferences about the target population.
  - Last (or next) birthday method.
    - Interviewer asks to speak to the person in the household who had the most recent birthday or who has the next birthday.
    - Not necessarily a probability sampling method because the people who first answer the survey request are more likely to say that they had most recent birthday or will have next birthday.

Appendix D

Additional information on different sampling techniques

Simple Random Sampling (SRS)

- SRS uses a sampling frame numbered 1 to *N* (the total number of elements on the frame). Random numbers from 1 to *N* are selected from a table of random numbers or a random number generator.

- Formula for estimating the sampling variance of a simple random sample:

  $\mathrm{var}(\bar{y}) = \frac{(1-f)s^2}{n}$, where

  f is the finite population correction and is equal to n (the sample size) divided by N (the number of elements on the sampling frame);
  s^2 is the sample element variance of the statistic of interest

  $$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

- The finite population correction indicates that, unlike the assumption made in standard statistical theory that the population is infinite, the survey population is finite in size and the sample is selected without replacement [16].

Systematic Sampling

- Steps of Systematic Sampling.
  - Compute the selection interval (*k*) as the ratio of the population size, *N*, to the sample size, *n*. In a formula, $k = \frac{N}{n}$.
  - Choose a random number from 1 to *k*.
  - Select the element of that random number from the frame and every *k*th element thereafter.

- Example 1.

- Imagine the size of the sampling frame is 10,000 and the sample size is 1,000, making the sampling interval, *k*, $\frac{10000}{1000}=10$. The sampler then selects a random number between 1 and 10, for instance, 6. The sampler will then make selections in this order – 6, 16, 26, 36…9996.
- Additional steps if the selection interval is a fraction:
    - Compute the selection numbers by adding the fractional sampling interval each time.
    - Drop the decimal portion of the selection numbers.

- Example 2.
    - The size of the sampling frame is 10,400 and the sample size is 1,000, making the sampling interval, *k*, $\frac{10400}{1000}=10.4$. The sampler selects a random number between 1 and 10.4, for instance, 6. The selection numbers would then be – 6, 16.4, 26.8, 37.2…10395.6. After rounding down, the selection numbers become – 6, 16, 26, 37…10395.

Stratified Sampling

- Stratified sampling steps:
    - Find information for every element on the frame that can be used to partition the elements into strata. Use information that is correlated to the measure(s) of interest. Each element on the frame can be placed in one and only one group.
    - Sort the frame by strata.
    - Compute a sample size (see Guideline 5).
    - Determine the number of sample selections in each respective stratum (allocation).

- There are 3 main types of allocation:
    - Proportionate allocation.
        - Selecting the sample so that elements within each stratum have the same probabilities of selection. Another way to conceive of proportionate allocations is that the sampler selects a sample of size n_h from each stratum *h* such that the proportion of elements in the sample from stratum *h,* $\frac{n_h}{n}$, is the same as the proportion of elements on the frame from stratum N_h, $\frac{N_h}{N}$.
    - Equal allocation.

- An allocation where the same number of elements are selected from each stratum.
- If one knows that all strata have equal distributions of the statistic of interest on the sampling frame, an equal allocation will create the highest level of precision in the sample estimate.

- Optimal allocation.
  - An allocation that produces the highest precision (i.e., narrowest confidence intervals) for the sample mean of any statistic of interest.
  - The sampler needs accurate estimates of the distributions of the frame elements for each stratum on the statistic of interest.

Cluster Sampling.

- Within-cluster homogeneity:
  - When selecting humans, it is important to consider that humans within a cluster tend to be more similar than humans across clusters because of:
    - Environment.
    - Self-selection.
    - Interaction with one another.
  - Since elements within a cluster tend to be alike, we receive less new information about the population when we select another element from that cluster rather than from another cluster. This lack of new information makes a cluster sample less precise than a stratified or even simple random sample. The rate of homogeneity (*roh*) is a way to measure this clustering effect.

Design effect

- A survey's design effect is defined as the ratio of the sampling variance under the complex design to the sampling variance computed as if a simple random sample of the same sample size had been selected. The purpose of the design effect is to evaluate the impact of the complex survey design on sampling variance measured to the variance of simple random sampling as the benchmark.
- For a cluster sample, the design effect is the effect of having chosen sampled clusters instead of elements. Due to within-cluster homogeneity, a clustered sample cannot assure representation of specified population subgroups as well as SRS, and will tend to have a design effect greater than one. On the other hand, stratification tends to generate design effects less than one since it ensures that specified population groups will be allocated at least one sample selection.

- In general, clustering increases the design effect, while stratification decreases it.
- Formulas:

Stratified designs

$$d_{eff} = \frac{\mathrm{var}(\bar{y}_{complex})}{\mathrm{var}(\bar{y}_{SRS})}$$

where d_{eff} is the design effect;

$\mathrm{var}(\bar{y}_{complex})$ is the variance of the complex sample design, whether it be stratfied only, clustered only, or a stratified cluster design;

$\mathrm{var}(\bar{y}_{SRS})$ is the variance of a SRS design, with the same sample size

$$d_{eff} = 1 + (b - 1)roh$$

where d_{eff} is the design effect;

b is the number of subselections within a selected cluster;

and roh is the rate of homogeniety

- In order to estimate the design effect for a new study, the roh is calculated from an earlier survey on a similar topic within a similar target population.

- Subsampling within selected clusters (multi-stage sampling).
  - $n = a*b$, where n is the sample size, a is the number of clusters selected and b is the number of selections within each cluster.
  - Pros: reduces the design effect and makes estimates more precise.
  - Cons: increases total costs because need to send interviewers to more areas.

Probabilities Proportional to Size (PPS)

- Situations where clusters are all of equal size rarely occur. PPS can control the sample size while ensuring that each element on the sampling frame has an equal chance of selection.

- Probabilities at either the first or second stage can be changed to ensure equal probabilities of selection for all elements.

- Imagine a two-stage cluster design where the clusters were blocks and the elements were housing units (HUs). The PPS formula would be:

$$f = f_{block} * f_{hu} = \frac{\alpha B_{\alpha}}{\sum B_{\alpha}} * \frac{b}{B_{\alpha}}$$

where
f is the overall probability of selection of the element,
f_{block} is the probability of selection of the cluster, and
f_{hu} is the probability of selection of the element within the cluster,
α is the number of cluster sections,
B_{α} is the number of elements within the selected sections α on the frame,
$\sum B_{\alpha}$ is the number of elements on the frame,
b is the number of elements selected within cluster α.

Example

Block #	Housing Units in Block	Cumulative Housing Units
1	25	25
2	30	55
3	35	90
4	40	130
5	20	150

- The sampler has the above list of blocks and wants to select three blocks (*a*), keep the sample size constant at 15 HUs and ensure that each HUs has the same probability of selection of one in ten (*f*=15/150). Using cumulative totals, numbers can be assigned to each block. Block 1 is assigned numbers 1-25, Block 2 26-55, Block 3 56-90, Block 4 91-130 and Block 5 131-150. From here, systematic sampling can be used to obtain a simple, without replacement sample of blocks based on the HUs within each block. Based on the frame size of 150 ($\sum B_{\alpha}$) and the number of selections being three, the selection interval is 50. Suppose the sampler chooses a random start of 29. In this case, the selection numbers would be 29, 79, and 129 corresponding to selections of Block 2, Block 3 and Block 4. To determine the selection probability of the HUs within Block 2 (f_{hu}), use the formula:

$$f = f_{block\,2} * f_{hu}$$

$$\frac{1}{10} = 3\left(\frac{30}{150}\right) * f_{hu}$$

$$f_{hu} = \frac{1}{10} * \frac{150}{90} = \frac{1}{6}$$

Since the selection probability of HUs within Block 2 is 1/6, the number of HUs selected within Block 2 (*b*) will be 30*1/6 or 5. Going through the same calculations for Blocks 3 and 4 will show that each block will have five selections.

- Potential problems and solutions with PPS sampling.
  - Problem: The same cluster may be chosen more than once.

    Solution: Use systematic selection with PPS [23].
  - Problem: Some of the clusters may not be large enough to produce subsamples of the required size.

    Solution: Link clusters to create new clusters that are all of sufficient size.
  - Problem: Some of the clusters are too large and the probability of selecting the cluster is greater than one.

    Solution: Remove the cluster from the list and choose elements from it directly.

Two-Phase Sampling

- Suggested steps [10]:
  - Phase 1 – Conduct a survey on a probability sample, using a relatively inexpensive data collection method subject to higher nonresponse rates than more expensive methods (see Data Collection).
  - Once the survey is completed, select a probability subsample of the nonrespondents to the Phase 1 survey.
  - Phase 2 – Use a more expensive method that generally produces lower nonresponse on the subsample.
  - Combine the results of the two surveys, with appropriate selection weights to account for unequal probabilities of selection between the selected respondents.

Panel Designs

- Three concerns about panel designs:

- The effort and costs of tracking and locating respondents who move over the duration of the panel survey.
- The change in the elements on the sampling frame over time. For example, in a cross-cultural panel survey of persons age 65 and older, some members of the original sampling frame will die, while other people will become eligible for selection.
- The repeated questioning of the same subjects over time may change how the subjects act and answer the questions (i.e., panel conditioning effect).

Appendix E

Sample Size Determination

- Recommended steps.
  - Define how many nested cells will be relevant for the analysis and what should be the minimal number of cases in each cell allowing for substantial analyses.
  - Have the survey sponsor specify the desired level of precision.
  - Convert these 95% confidence intervals into a sampling variance of the mean, $\text{var}(\bar{y})$.
    - Example: The survey sponsor wants a 95% confidence interval of .08 around the statistic of interest. Since the half width of a 95% confidence interval (CI) is $\frac{1}{2}(95\%\ CI) = 1.96(se(\bar{y}))$. This formula can be rearranged with basic algebra to calculate the precision (sampling variance of the mean) from this confidence interval:

$$\text{var}(\bar{y}) = (se(\bar{y}))^2 = \left(\frac{.5(95\%\ \text{Conf. Interval})}{1.96}\right)^2 = \left(\frac{.04}{1.96}\right)^2 = .0004165.$$

  - Obtain an estimate of S^2 (population element variance).
  - If the statistic of interest is not a proportion find an estimate of S^2 from a previous survey on the same target population or a small pilot test.
  - If the statistic of interest is a proportion, the sampler can use the expected value of the proportion (*p*), even if it is a guess, to estimate S^2 by using the formula $s^2 = p(1-p)$.
  - Estimate the needed number of completed interviews for a simple random sample (SRS) by dividing the estimate of S^2 by the sampling variance of the mean.
    - Example: the obtained estimate of S^2 is .6247. Therefore the needed number of completed interviews for an SRS (n_{srs}) is:

$$n_{srs} = \frac{.6247}{.0004165} = 1,499.88 \approx 1,500.$$

  - Multiply the number of completed interviews by the design effect to account for a non- SRS design.
    - Example: the design effect of a stratified clustered sample is 1.25. Taking into account the design effect, the number of completed interviews for this complex (i.e., stratified clustered) sample is: $n_{complex} = n_{srs} * d_{eff} = 1,500 * 1.25 = 1,875$.

- The sample size must account for three additional factors:
- Not all sampled elements will want to participate in the survey (i.e., response rate).
- Not all sampled elements, given the target population, will be eligible to participate (i.e., eligibility rate).
- The frame will likely fail to cover all elements in the survey population (i.e., coverage rate).
- Calculate the necessary sample size by dividing the number of completed interviews by the expected response rate, eligibility rate, and coverage rate.
- The sampler can estimate these three rates by looking at the rates obtained in previous surveys with the same survey population and survey design.
  - Example: The expected response rate is 75%, the expected eligibility rate is 90%, and the expected coverage rate is 95%. Therefore, the necessary sample size is:

$$n_{final} = \frac{n_{complex}}{Resp\ rate * Elig\ rate * Cov\ rate} = \frac{1875}{.75 * .9 * .95} = 2923.97 \approx 2924.$$

Glossary

Accuracy	The degree of closeness an estimate has to the true value.
Audit Trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.
Auxiliary data	Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in imputation techniques.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Cluster sampling	A sampling procedure where units of the sampling frame that are similar on one or more variables (typically geographic) are organized into larger groups (i.e. clusters), and a sample of groups is selected. The selected groups contain the units to be included in the sample. The sample may include all units in the selected clusters or a sub-sample of units in each selected cluster. The ultimate purpose of this procedure is to reduce interviewer travel costs for in person studies by producing distinct groups of elements where the elements within each group area are geographically close to one another.

Codebook	A document that provides question-level metadata that is matched to variables in a dataset. Metadata include the elements of a data dictionary, as well as basic study documentation, question text, universe statements (the characteristics of respondents who were asked the question), the number of respondents who answered the question, and response frequencies or statistics.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Complex survey data (or designs)	Survey datasets (or designs) based on stratified single or multistage samples with survey weights designed to compensate for unequal probabilities of selection or nonresponse.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Convenience sample	A sample of elements that are selected because it is convenient to use them, not because they are representative of the target population.
Cooperation rate	The proportion of all elements interviewed of all eligible units ever contacted.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Coverage	The proportion of the target population that is accounted for on the sampling frame.

Coverage error	Survey error (variance and bias) that is introduced when there is not a one-to-one correspondence between frame and target population units. Some units in the target population are not included on the sampling frame (undercoverage), some units on the sampling frame are not members of the target population (out-of-scope), more than one unit on the sampling frame corresponds to the same target population unit (overcoverage), and one sampling frame unit corresponds to more than one target population unit.
Coverage rate	The number of elements on the sampling frame divided by the estimated number of elements in the target population.
Data dictionary	A document linking the survey instrument (questionnaire) with the dataset, or more abstract question or variable-level metadata including question identifiers (variable names and labels); response category identifiers (value labels), and data types (e.g., F2.0, specifying that the response is a two-digit integer with zero decimal places.
Design effect	The effect of the complex survey design on sampling variance measured as the ratio of the sampling variance under the complex design to the sampling variance computed as a simple random sample of the same sample size.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Eligibility rate	The number of eligible sample elements divided by the total number of elements on the sampling frame.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."

Fixed panel design	A longitudinal study which attempts to collect survey data on the same sample elements at intervals over a period of time. After the initial sample selection, no additions to the sample are made.
Fixed panel plus births design	A longitudinal study in which a panel of individuals is interviewed at intervals over a period of time and additional elements are added to the sample.
Half open interval	A method of updating lists of addresses by adding previously omitted units to the sample when the units are identified geographically next to a selected unit.
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Informant	The person who supplies a list of the eligible elements within the selected unit. For example, many in-person surveys select a sample of housing units at the penultimate stage of selection. Interviewers then contact the housing unit with the aim of convincing the member of the housing unit who responded to the contact attempt to provide a list of housing unit members who are eligible for the study. The housing unit member who provides a list of all eligible housing unit members is called the informant. Informants can also be selected respondents as well, if they are eligible for the study and are chosen as the respondent during the within household stage of selection.
Interpenetrated sample assignment, interpenetration	Randomized assignment of interviewers to subsamples of respondents in order to measure correlated response variance, arising from the fact that response errors of persons interviewed by the same interviewer may be correlated. Interpenetration allows researchers to disentangle the effects interviewers have on respondents from the true differences between respondents.
Interviewer variance	That component of overall variability in survey estimates that can be accounted for by the interviewers.

Item nonresponse, item missing data	The lack of information on individual data items for a sample element where other data items were successfully obtained.
Listing	A procedure used in area probability sample designs to create a complete list of all elements or cluster of elements within a specific set of geographic boundaries.
Longitudinal study	A study where elements are repeatedly measured over time.
Majority country	A country with low per capita income (the majority of countries).
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Metadata	Information that describes data. The term encompasses a broad spectrum of information about the survey, from study title to sample design, details such as interviewer briefing notes, contextual data and/or information such as legal regulations, customs, and economic indicators. Note that the term 'data' is used here in a technical definition. Typically metadata are descriptive information and data are the numerical values described.
Mode	Method of data collection.
Noncontact	Sampling units that were potentially eligible but could not be reached.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Nonresponse bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.

Outcome rate	A rate calculated based on the study's defined final disposition codes that reflect the outcome of specific contact attempts before the unit was finalized. Examples include response rates (the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.), cooperation rates (the proportion of all units interviewed of all eligible units ever contacted), refusal rates (the proportion of all units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units), and contact rates (the proportion of all units are reached by the survey).
Paradata	Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.
Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins.
Poststratification	A statistical adjustment that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection.
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.

Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Probability proportional to size (PPS)	A sampling method that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for postratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection.
Probability sampling	A sampling method where each element on the sampling frame has a known, non-zero chance of selection.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.

Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Quota sampling	A non-probability sampling method that sets specific sample size quotas or target sample sizes for subclasses of the target population. The sample quotas are generally based on simple demographic characteristics, (e.g., quotas for gender, age groups and geographic region subclasses).
Random-digit-dialing (RDD)	A method of selecting telephone numbers in which the target population consists of all possible telephone numbers, and all telephone numbers have an equal probability of selection.
Refusal rate	The proportion of all units of all potentially eligible sampling units in which a respondent sampling unit refuses to do an interview or breaks off interviews of all potentially eligible sampling units.
Repeated panel design	A series of fixed panel surveys that may or may not overlap in time. Generally, each panel is designed to represent the same target population definition applied at a different point in time.
Replicates	Systematic probability subsamples of the full sample.
Residency rule	A rule to help interviewers determine which persons to include in the household listing, based on what the informant reports.
Response options	The category, wording, and order of options given with the survey question.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.

Rotating panel design	A study where elements are repeatedly measured a set number of times, then replaced by new randomly chosen elements. Typically, the newly-chosen elements are also measured repeatedly for the appropriate number of times.
Sample design	Information on the target and final sample sizes, strata definitions and the sample selection methodology.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sampling error	Survey error (variance and bias) due to observing a sample of the population rather than the entire population.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Sampling variance	A measure of how much a statistic varies around its mean (over all conceptual trials) as a result of the sample design only. This measure does not account for other sources of variable error such as coverage and nonresponse.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.

Simple random sampling (SRS)	A procedure where a sample of size n is drawn from a population of size N in such a way that every possible sample of size n has the same probability of being selected.
Split panel design	A design that contains a blend of cross-sectional and panel samples at each new wave of data collection.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Stratification	A sampling procedure that divides the sampling frame into mutually exclusive and exhaustive groups (or strata) and places each element on the frame into one of the groups. Independent selections are then made from each stratum, one by one, to ensure representation of each subgroup on the frame in the sample.
Substitution	A technique where each nonresponding sample element from the initial sample is replaced by another element of the target population, typically not an element selected in the initial sample. Substitution increases the nonresponse rate and most likely the nonresponse bias.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Survey weight	A statistical adjustment created to compensate for complex survey designs with features including, but not limited to, unequal likelihoods of selection, differences in response rates across key subgroups, and deviations from distributions on critical variables found in the target population from external sources, such as a national Census.
Systematic sampling	A procedure that selects of every kth element on the sampling frame after a random start.

Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Tracking	The process of attempting to locate a sample element that changed contact information (e.g. address, telephone number, email address) since the last time the element's contact information was collected.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Universe statement	A description of the subgroup of respondents to which the survey item applies (e.g., "Female, ≥ 45, Now Working").
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

References

[1] Barahona, C., & Levy, S. (2006). The best of both worlds: Producing national statistics using participatory methods. *World Development, 35*(2), 326-341.

[2] Bergsten, J. W. (1980). *Some sample survey designs in Syria, Nepal and Somalia.* Paper presented at the Proceedings of American Statistical Association: Survey Research Methods Section. Retrieved March 25, 2010, from http://www.amstat.org/Sections/Srms/Proceedings/papers/1980_013.pdf

[3] Binder, D. (1998). Longitudinal surveys: Why are these surveys different from all other surveys? *Survey Methodology*, *24*(2), 101-108.

[4] Chikwanha, A. B. (2005). *Conducting surveys and quality control in Africa: Insights from the Afrobarometer*. Paper presented at the WAPOR/ISSC Conference, Ljubljana, Slovenia.

[5] China Health and Retirement Longitudinal Study (CHRLS). (2009). *Advance team manual.* Retrieved November 15, 2009 from http://charls.ccer.edu.cn/charls/documentation.asp

[6] Chinese Family Panel Studies (CFPS). (2010). *Listing manual (draft)*. Unpublished manuscript.

[7] Cochran, W. G. (1977). *Sampling techniques*. New York, NY: John Wiley & Sons.

[8] Dunzhu, S., Wang, F. S., Courtright, P., Liu, L., Tenzing, C., Noertjojo, K., et al. (2003). Blindness and eye diseases in Tibet: Findings from a randomised, population based survey. *Br J Ophthalmol, 87*(12), 1443-1448.

[9] Groves, R. M. (1989). *Survey errors and survey costs*. New York, NY: John Wiley & Sons.

[10] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[11] Groves, R. M. & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*, 439–457.

[12] Groves, R. M., & Lepkowski, J. M. (1985). Dual frame, mixed mode survey designs. *Journal of Official Statistics, 1*(3), 263-286.

[13] Groves, R. M., Kirgis, N., Peytcheva, E., Wagner, J., Axinn, W., & Mosher, W. (2008). *Responsive design for household surveys: Illustration of management interventions based on survey*

paradata. Paper presented at the 2008 AAPOR meetings.

[14] Häder, S., & Gabler, S. (2003). Sampling and estimation. In J. A. Harkness, F. J. R. van de Vijver & P. M. Mohler (Eds.), *Cross-cultural survey methods,* (pp. 117-136). Hoboken, NJ: John Wiley & Sons.

[15] Hartley, H. O. (1962). Multiple-frame surveys. *Proceedings of the social statistics section* (pp. 203-206). Alexandria, VA: American Statistical Association.

[16] Heeringa, S. G., & O'Muircheartaigh, C. (2010). Sampling design for cross-cultural and cross-national studies. In J. Harkness, B. Edwards, M. Braun, T. Johnson, L. E. Lyberg, P. Mohler, B-E. Pennell, & T. Smith. (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts* (251-267). Hoboken: John Wiley & Sons.

[17] Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2008). *Private household: Concepts and their operationalization in national and international social science surveys* (In German). Retrieved February 13, 2010, from http://www.ssoar.info/ssoar/View/?resid=2660

[18] Kalton, G. (1983). *Introduction to survey sampling.* Newbury Park, CA: Sage Publications.

[19] Kasprzyk, D. (1988). *The survey of income and program participation: An overview and discussion of research issues* (No. 77). Washington, DC: U.S. Bureau of the Census. Retrieved March 26, 2010, from http://www2.census.gov/prod2/sipp/wp/SIPP_WP_077.pdf

[20] Kasprzyk, D., Duncan, G.J., Kalton, G., Singh, M.P. (Eds.). (1989). *Panel surveys*. New York, NY: John Wiley & Sons.

[21] Kessler, R. C., Ustun, T. B., & World Health Organization. (2008). *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge; New York: Geneva: Cambridge University Press; Published in collaboration with the World Health Organization.

[22] Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, *44*, 380-387.

[23] Kish, L. (1965). *Survey sampling.* New York, NY: John Wiley & Sons.

[24] Kish, L. (1987). *Statistical design for research*. New York, NY: John Wiley & Sons.

[25] Kish, L. (1994). Multipopulation survey designs: Five types with seven

shared aspects. *International Statistical Review, 62*, 167–186.

[26] Lavallée, P., Michaud, S., & Webber, M. (1993). The survey of labour and income dynamics, design issues for a new longitudinal survey in Canada. *Bulletin of the International Statistical Institute* (49th Session, Book 2) (pp. 99-100).

[27] Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

[28] Lynn, P. (Ed.) (2009). *Methodology of longitudinal surveys*. New York, NY: John Wiley & Sons.

[29] Lynn, P., Häder, S., Gabler, S., & Laaksonen, S. (2007). Methods for achieving equivalence of samples in cross-national surveys: The European Social Survey experience. *Journal of Official Statistics, 23*(1), 107–124.

[30] Mankoff, C. (2009). *Assisting the survey listing process using geography*. [PowerPoint slides]. Retrieved November 15, 2009, from http://www.cpc.unc.edu/urbanreproductivehealth/tools-and-resources/workshop-materials/Assisting_the_Survey_Listing_Process_Web.pdf

[31] Martin, E. (1999). Who knows who lives here? Within-household disagreements as a source of survey coverage error. *Public Opinion Quarterly, 63*(2), 220-236.

[32] Okafor, R., Adeleke, I., & Oparac, A. (2007). *An appraisal of the conduct and provisional results of the Nigerian Population and Housing Census of 2006*. Paper presented at the Proceedings of American Statistical Association: Survey Research Methods Section. Retrieved March 26, 2010, from http://www.amstat.org/Sections/Srms/Proceedings/y2007/Files/JSM2007-000561.pdf

[33] Peracchi, F. (2002). The European community household panel: A review. *Empirical Economics*, *27*, 63-90.

[34] Scholz, E. & Heller, M. (2009). *ISSP Study Monitoring 2007* (GESIS Technical Reports, 2009/5). Mannheim: GESIS. Retrieved March 26, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/TechnicalReport_09-

5.pdf

[35] Survey Research Center. (2006). *Listing procedures manual*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

[36] Survey Research Center. (1990). *National area sample sampling manual-Listing instructions.* Ann Arbor, MI: Institute for Social Research, University of Michigan.

[37] Survey Research Center. (1985). *Household survey of Grenada*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

[38] Tourangeau, R., Shapiro, G., Kearney, A., & Ernst, L. (1997). Who lives here? Survey undercoverage and household roster questions. *Journal of Official Statistics, 13*, 1-18.

[39] Tucker, C., Lepkowski, J. M., & Piekarski, L. (2002). The current efficiency of list assisted telephone sampling designs. *Public Opinion Quarterly*, *66*(3), 321-338.

[40] United Nations. (1998). *Principles and recommendations for population and housing censuses, revision 1* (Statistical Papers, No. 67/Rev. 1, Sales No. E.98.XVII.8). New York, NY: United Nations. Retrieved March 26, 2010, from http://unstats.un.org/unsd/pubs/gesgrid.asp?ID=127

[41] Üstun, T. B., Chatterji, S., Mechbal, A., & Murray, C. J. L. (2005). Chapter X: Quality assurance in surveys: Standards, guidelines, and procedures. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), *Household surveys in developing and transition countries*. New York, NY: United Nations. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[42] World Values Survey (WVS). *Mission statement and goals of the World Values Survey Association*. Retrieved July 10, 2010, from http://www.worldvaluessurvey.org/wvs/articles/folder_published/article_base_52/files/constitution.pdf

[43] Yansaneh, I. (2005). Chapter II: Overview of sample design issues for household surveys in developing and transition countries. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), *Household surveys in developing and transition countries.* New York, NY: United Nations. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

Internet Links

[44] Afrobarometer Survey. Retrieved July 1, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[45] Asian Barometer. Retrieved July 1, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

[46] European Social Survey (ESS). (2010). Retrieved May 03, 2010, from http://www.europeansocialsurvey.org

[47] European Social Survey (ESS) - Sample Design. Retrieved July 10, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=80&Itemid=365

[48] International Social Survey Programme. Retrieved July 10, 2010, from http://www.issp.org/

[49] Survey of Health, Ageing and Retirement in Europe (SHARE). (2010). Retrieved July 10, 2010, from http://www.share-project.org/

[50] World Mental Health Study. (2010). Retrieved May 04, 2010, from http://www.hcp.med.harvard.edu/wmh/

[51] World Values Survey. Retrieved July 10, 2010, from http://www.worldvaluessurvey.com/

Further Reading

<u>Sampling</u>

Cochran, W. G. (1977). *Sampling techniques*. New York, NY: John Wiley & Sons.

Kalton, G. (1983). *Introduction to survey sampling.* Newbury Park, CA: Sage Publications.

Kish, L. (1965). *Survey sampling.* New York, NY: John Wiley & Sons.

Lohr, S. (2009). *Sampling: Design and analysis* (2nd ed.). Pacific Grove, CA: Duxbury Press.

<u>Statistics</u>

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Ames, IA: Iowa State University Press.

VI. Questionnaire Design

Janet Harkness with (alphabetically) Ipek Bilgen, AnaLucía Córdova Cazar, Lei Huang, Debbie Miller, Mathew Stange, and Ana Villar

Introduction

The following guidelines present options for the deliberate design of questions intended for implementation in multiple cultures and languages. In this context, “deliberate design” means that the questions have been specifically constructed or chosen for comparative research purposes, according to any of several criteria and strategies [12]. The models and strategies discussed here are applicable to a variety of disciplines, including the social and behavioral sciences, health research, and attitudinal survey research.

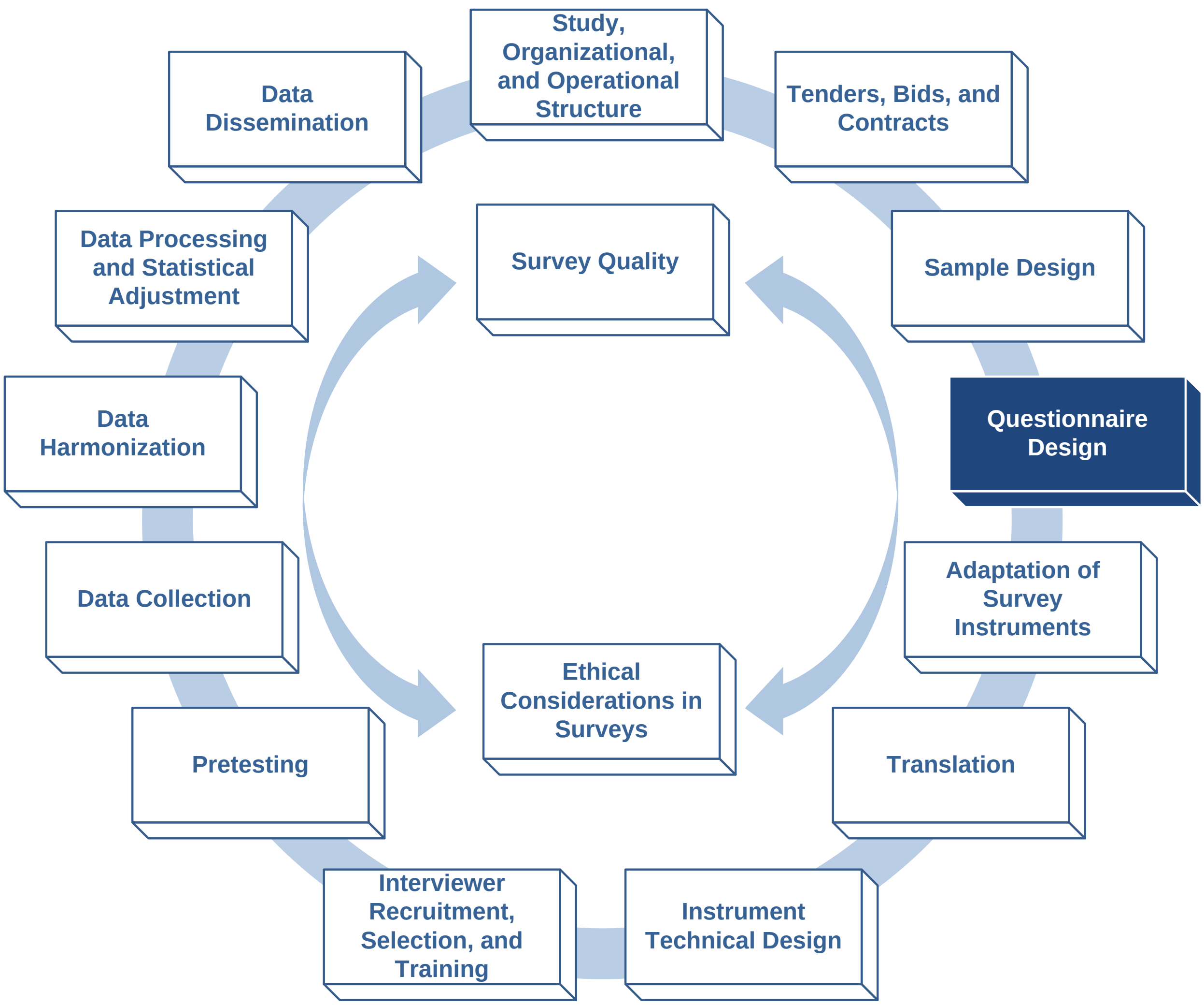
This chapter presents a basic outline of the approaches available to develop questions for comparative studies, the procedures involved in each, and the advantages and disadvantages of the different approaches.

Although comparative questionnaire design is related in various ways to question translation, adaptation, technical instrument design, pretesting, and harmonization, these topics are more fully addressed in other chapters (see Translation, Adaptation of Survey Instruments, Instrument Technical Design, Pretesting, and Translation: Language Harmonization).

This chapter borrows terminology from translation studies, which define “source language” as the language translated out of and “target language” as the language translated into. In like fashion, the chapter distinguishes between “source questionnaires” and “target questionnaires.” Source questionnaires are questionnaires used as a blueprint to produce other questionnaires, usually on the basis of translation into other languages (see Translation); target questionnaires are versions produced from the source questionnaire, usually on the basis of translation or translation and adaptation (see Adaptation of Survey Instruments). Target questionnaires enable researchers to study populations who could not be studied using the source questionnaire.

Figure 1 shows questionnaire design within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



This chapter begins with a brief review of topics relevant for sound questionnaire design in general, leaving aside any comparative considerations. Those with extensive experience in designing questionnaires may choose to skip Guideline 1.

Guidelines

Goal: To maximize the comparability of survey questions across cultures and languages and reduce measurement error related to question design.

1. **Ensure that questionnaire design follows basic best practice recommendations for general survey research, whenever and wherever these are appropriate for a given comparative context.**

Rationale

Basic questionnaire design requirements need to be met in any project, whether comparative or not.

The procedural steps presented here identify fundamental aspects of questionnaire design with which researchers should be familiar before beginning work on any questionnaire and certainly before attempting comparative design. The steps do not provide guidance on each facet of design identified or on general design issues. A wealth of survey literature addresses these topics [4] [5] [8] [10] [28].

Procedural steps

- Review survey methods literature and research on basic aspects of general questionnaire design. Theories contributing to question/questionnaire design include:
  - Cognition and survey research, including theories of survey response [22] [24].
  - Response styles and bias [1] [17] [23] [29].
  - Pragmatic aspects of the survey interview event [22].

- Review literature and research on the craft, technique, and art of asking questions [4] [5] [7] [8] [10] [21]. Specific areas to consider are kinds of questions, question formats, response scales and mode.
  - Kinds of questions (some of the distinctions made below overlap; for example, a factual question may be about behavior or may ask for socio-demographic details):
    - Factual questions. Factual questions aim to collect information about things for which there is a correct answer. In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative

records. Factual questions can be about a variety of things, such as figure-based facts (date, age, weight), events (pregnancy, marriage), and behaviors (smoking or media consumption).

Example: *Do you smoke?*

- Socio-demographic questions. Socio-demographic questions typically ask about respondent characteristics such as age, marital status, income, employment status, and education. They can be intended for use as either dependent or independent variables. For discussion of their design and interpretation in the comparative context, see [9] [15] [16]. See also Translation: Language Harmonization and Adaptation of Survey Instruments.

  Example: *What year and month were you born?*
- Behavioral questions. Behavioral questions ask people to report on things they do or have done.

  Example: *Have you ever smoked cigarettes?*
- Attitudinal questions. Attitudinal questions ask about respondents' opinions, judgments, emotions, and perceptions. These cannot be measured by other means; we are dependent on respondents' answers.

  Example: *Do you think smoking cigarettes is bad for the smoker's health?*
- Intention questions on behavior. Intention questions ask respondents to indicate their intention regarding some behavior. They share features with attitudinal questions.

  Example: *Do you intend to stop smoking?*

▪ Question formats:

- Closed question format. In closed question formats, the survey question provides a limited set of predefined answer categories from which respondents choose.

  Example: *Do you smoke?*

  Yes ___
  No ___
- Open question format. Open question formats require respondents to answer questions in their own words.

  Example: *What is your occupation?*
  (Please write in the name or title of your occupation_______)

▪ Response scales are predefined sets of answer categories for a closed question from which respondents are asked to select a response. Common response scale formats are rating and ranking formats and frequency scales.

- Rating. Rating is a response format which requires the respondent to select one position on an ordered scale of answer options.

Example: *To what extent do you agree or disagree with the following statement?*
It is a good idea to ban smoking in public places.
Strongly agree
Somewhat agree
Neither agree nor disagree
Somewhat disagree
Strongly disagree

- Ranking. Ranking is a response format where respondents express their preferences by ordering persons, brands, etc. from top to bottom, generating a rank order of a list of items or entities.

  Example*: Listed below are possible disadvantages related to smoking cigarettes. Please enter the number 1, 2, 3, or 4 alongside each possible disadvantage to indicate your rank ordering of these. 1 stands for the greatest disadvantage, 4 for the least disadvantage.*
  ______ *Harmful effects on other people's health*
  ______ *Stale smoke smell in clothes and furnishings*
  ______ *Expense of buying cigarettes*
  ______ *Harmful effects on smoker's health*

▪ Mode [6] [7] (i.e., the means by which data are collected). The choice of mode will affect the design options for various aspects of questionnaire and survey instrument design (e.g., length of the questionnaire, layout of instruments, and application of visual stimuli). (See Instrument Technical Design.)
- In terms of the standard literature, "mode" is related to whether an interviewer enters the data (as in telephone and face-to-face interviews) or the respondent enters the data (as in web surveys and paper-and-pencil mail surveys).
- A second relevant aspect is the channel of communication (visual, oral, aural, tactile).
- A third is the locus of control (who, for instance, determines what is presented and in what order).

2. Become familiar with the comparative design options available and the advantages and disadvantages of each.

Rationale

Knowledge of the different approaches available for comparative design for surveys in multiple cultures, languages, or countries enables researchers to make informed choices for their projects.

Procedural steps

- Read relevant literature (and, if possible, talk to primary researchers) to become familiar with the advantages and disadvantages of the major approaches. The three basic approaches involve asking the same questions (ASQ), asking different questions (ADQ), or using a mixed approach that combines ASQ and ADQ [11] [12] [14].
  - Ask the same questions (ASQ). In this approach to question design, researchers ask a common set of questions of all populations studied.
    - The most common way to do this is by developing a source questionnaire in one language and then producing other language versions, usually on the basis of translation or translation and adaptation.
    - Decentering is a second way to "ask the same questions". However, decentering is basically only suitable for two language projects under special circumstances [11] [12] [27].
  - Ask different questions (ADQ). In this approach, researchers ask the most salient questions for each population on a given common construct or conceptual domain. The different questions and, possibly, different indicators used in each location are assumed to tap a construct that is shared across populations.
  - A mixed approach that combines ASQ and ADQ.

- Weigh the advantages and disadvantages of each approach in terms of the study design (see overview in Appendix A).

- Decide on the most viable approach for the study within a quality framework that addresses survey error related to questionnaire design. (see Survey Quality).

Lessons learned

- Not all options will be available for every study. The study design, the target population, and the mode required may all impose constraints. For example, if questions from other studies are to be used again ("replicated"), only an ASQ model based on translation (perhaps with adaptation) is possible for these questions. The chosen data collection method, the sample design, the fielding schedules, and available funds or stipulations on the use of funds can all limit options (see Data Collection, Sample Design, and Tenders, Bids, and Contracts).

- Literature explaining options can be hard to locate, unclear, or very sparse on details. Even detailed study reports might be clear to people involved in a project but not clear enough for "outside" readers.

- Researchers are usually most familiar with the ASQ approach based on translation, but may not be aware of the limitations and constraints of this approach [2] [12] [13] [14]. In addition, pressures to replicate questions might over-promote the ASQ-and-translate (ASQT) approach.

- Comparability or “equivalence” is sometimes judged on the basis of similar wording across questionnaires. This is, indeed, what is often targeted in ASQT approaches. However, even nominally "accurate" translations do not necessarily produce comparable data (see Translation). For example, a close translation of the English question “Does he like adventures?” in French is more likely to be understood as “Does he like amorous adventures?”

- Most cross-cultural and cross-national surveys use a mix of ASQ and ADQ questions. Some questions will also be a blend of a common part (ASQ) and country-specific parts (ADQ). Socio-demographic questions on education, for example, are often asked in terms of a shared question stem (such as "What is the highest level of education you have completed?"), accompanied by local/national categories of educational level or qualification (ADQ). These are then mapped onto an international standard (see Language Harmonization).

- It is difficult to find examples of surveys with most substantive questions based on an ADQ approach. There are examples of research that analyzes different questions from different studies and takes them to reflect aspects of a given common construct [26].

3. **Establish a lead team or working group responsible for questionnaire design, and appoint a coordinator responsible for organizing scheduling, communication channels and rules, and the design deliverables.**

Rationale

Good questionnaires can rarely be developed by a single person. This is especially true for cross-national or cross-cultural research. In accordance with a quality assurance framework for design, a team is needed that provides the spread of knowledge, diverse skills, and cultural backgrounds for which successful comparative design calls [19]. (See also Study, Organizational, and Operational Structure.)

Procedural Steps

- Decide, as appropriate, on the shared language(s) (lingua franca) and communication mediums to be used in the overall project and in the work of the questionnaire design team.

- Identify a lead person in the design team who is also responsible for coordinating with other groups in the project (such as the central coordinating center, if one exists).

- Identify the various skills required in the team.
  - These include all the skills needed for questionnaire design in general.
  - They also include special expertise or skills relevant for designing a comparative instrument (e.g., understanding approach options, understanding the cultural impact of conceptual coverage, common ground, comprehension, response styles, local population structure and needs, etc.).
  - Depending on their roles in the team, members may need to be conversant in whatever lingua franca is used in a multilingual project.

- Ensure that the members recruited are from a reasonable spread of the countries, locations, or cultures participating in the study.

- Ensure that the members recruited for the questionnaire design team have the skills and abilities needed for good questionnaire design. A questionnaire design team should consist of a mix of people bringing together good writing skills, measurement and statistical knowledge, cognitive and survey research expertise, information technology knowledge, and documentation know-how [20].

- Identify the responsibilities of each member at an appropriate level of detail.

- Recruit collaborators and external experts, as necessary and feasible, from the different populations involved. This ensures the availability of expertise on given topics and local knowledge. A drafting team might need specific and short-term input from an expert on a substantive area in the questionnaire. For example, if input on pensions is needed, an expert on the topic may be brought in exclusively for the development of pension-specific questions.

- Establish a community of cooperation and trust.

Lessons learned

- Ways should be found to have any groups who are participating in the project, but are not directly part of the core development team, to contribute to the development of the questionnaire.

4. **Establish the procedures and protocols for questionnaire development and for testing at different stages in this development.**
Rationale

Clear identification of the procedures and the protocols to be followed is essential to inform all those involved and to effectively implement and assess the chosen design process.

While different studies follow different design models (ASQ, ADQ, mixed approaches), this guideline identifies some of the key generic elements to be considered.

Procedural steps

- Establish which design and related procedures are to be used (e.g., ASQ source questionnaire and translation).
- Develop the protocols relevant for the chosen design model and the processes it calls for (e.g., protocol for questionnaire development of a source questionnaire intended for use in multiple locations and cultures/languages).
- Create a schedule and budget for the milestones, deliverables, and procedures involved in the chosen design model. In the ASQT model this would include schedules and a budget for producing draft source questionnaires, review by participating cultures or groups, deadlines for feedback, schedules for pretesting, schedules for language harmonization, schedules for translation, and subsequent assessment and pretesting.
- Create a framework of quality assurance and quality control to ensure compliance with protocols and the adequacy of outputs.
- Create communication channels and encouragements which ensure that participants can and do make feedback on draft designs they are asked to review.

Lessons learned

- Not all participating groups in a project will be confident that their input in the developmental process is (a) valuable in generic terms for the entire project, (b) accurate or justified, and (c) welcomed by perceived leading figures or countries in either the design team or the larger project. It is important to make clear to participating groups that every contribution is valuable. Sharing feedback across the project underscores the value of every contribution and explains to participating groups why their suggestions are or are not incorporated in design modifications.

5. Pretest source and target questionnaires.

Rationale

Questionnaires need to be pretested before they are used. The source questionnaire needs to be assessed for its suitability as a source questionnaire for multiple other versions, rather than as a questionnaire for a single population. The other versions produced—most likely on the basis of translation or translation and adaptation—also need to be pretested for suitability with a given target population.

Procedural steps

(For detailed information about pretesting, see Pretesting.)

Lessons learned

- Pretesting is essential. Even questions previously used in other questionnaires must be tested for their suitability in a new context and for use with new populations.

- Where possible, pretesting of the source questionnaire should be combined with pretesting a spread of other languages representing the diverse target populations in the project.

- Ensuring the quality of questionnaire development prior to pretesting is just as important as pretesting itself. Proper team selection, adequate briefing on requirements and expectations, and good use of documentation will enhance the quality of the questions presented for pretesting so that pretesting serves the monitoring and refining purposes it should have.

6. Establish a quality assurance and quality monitoring framework for questionnaire development.

Rationale

Irrespective of the design approach followed to produce a questionnaire, quality standards must be set. These are critical to establishing quality assurance and quality monitoring steps for the process of developing any questionnaire [16].

Procedural steps

- Be cognizant of possible sources of survey error in the questionnaire design phase and develop quality assurance and quality monitoring steps to address these. Possible sources of error in this phase include validity and measurement issues [10].
- Acquaint question designers with important quality assurance literature on the topic of question design (e.g., on validity, tests of conceptual coverage, response process, sources of measurement error) [3] [10].
- For source questionnaires, form a team in each country or location that meets to discuss the development and assessment of the source questionnaire at each phase. The team should have, or should be provided with, the methodological expertise needed for this task.
- Have such teams document and report any queries or problems to the questionnaire drafting group in a timely fashion during the development phases or, as appropriate, report to the project coordinating center [19].

Lessons learned

- Quality assurance and quality monitoring should be addressed early in the design planning process.
- Variations in country-level assessment experience, research traditions, and methodological rigor regarding question design need to be thoroughly investigated and understood when setting quality standards. Some locations or countries will need more assistance than others in understanding the relevance of some requirements. They may also need guidance on how products can be assessed in terms of these requirements.

- Some entity, such as a drafting group coordinator or a coordinating center, must be appointed to lead on these matters.

- Through their knowledge of their own location and culture, local level collaborators and team members may well provide insights that other team members lack, even if quite experienced in questionnaire design.

7. Develop qualitative and quantitative protocols and procedures for assessing the quality of questions across survey implementations.

Rationale

Identifying standards to be met and establishing the criteria required to meet them, as well as agreeing on the good/best practice procedures to follow, are basic to undertaking quality assurance and quality monitoring.

Procedural steps

- Determine appropriate methods to assess the quality of questions. Consider question standards and survey determinants (e.g., funding and resources), as well as the model of design chosen for the topic [10].

- Include qualitative and quantitative methods of assessment (see the Pretesting chapter for a detailed description of assessment methods).
  - Qualitative options include:
    - Various pretesting techniques, such as focus groups, cognitive testing, pilot studies, and prototype studies (see Pretesting).
    - Expert appraisals by such groups as target population members, substantive experts, question design experts, or translators.
    - Debriefings from any testing (interviewers and respondents).
  - Quantitative methods of assessment include [3] [25]:
    - Reliability (e.g., Cronbach's alpha) and validity.
    - Exploratory and confirmatory analyses such as multi-trait multi-method (MTMM), item response theory (IRT), differential item functioning (dif), variance analysis, factor analysis, or stand-alone or embedded experiments.

Lessons learned

- Both qualitative and quantitative methods of assessment are necessary. Reliance on one without the other is not advised.

- Do not use pretesting as the main tool for question refinement. Make the questions as well designed as possible before pretesting so that pretesting can be used to find problems not identifiable in other refinement procedures.

- Different disciplines favor and can use different developmental and testing procedures, partly because of their different typical design formats. Social Science surveys, for example, often have only one or two questions on a particular domain; psychological and educational scales, on the other hand, might have more than twenty questions on one domain.

8. Develop a documentation scheme for questionnaire design decisions, design implementation, and quality assurance protocols.

Rationale

Documentation aids in producing the questionnaire and can be a tool for quality assurance and monitoring. As the Survey Quality chapter indicates, continual measurement and documentation of the quality targeted and achieved is necessary to identify quality problems. Even if sources of error are not recognized until later, documentation can be used to inform improved designs in future studies.

Procedural steps

- Design the documentation process before question development begins and document question design from the start. This ensures that all decisions are captured and that action can be taken in a timely fashion.

- Standardize documentation requirements and formats across all locations or countries involved in question development. This facilitates feedback in an ASQ model and comparable development in an ADQ model.

- Create flexible documentation templates that allow categories to be added if unforeseen issues arise.

- Create a clear and concise description of the questionnaire design procedures which is user-oriented and user-friendly. Include:
  - Conceptualization from concept to questions.
  - Operationalization (approach, mode, development across versions, adaptation agreements, annotations, language harmonization, origin of questions whether new, replicated, adopted, or adapted).

    - Analysis plan.

- Record the development of indicators and questions from start to finish (e.g., any modifications made to questions at different stages and why).

- Version control procedures could be necessary if a source questionnaire is modified across time.
    - A version of the source questionnaire will serve as the gold standard, or source version 1. Document any changes made to it over time.

Lessons learned

- Documentation must accompany questionnaire design since it will be used to detect problems in time to address them.

- If documentation is left to the end of questionnaire design (or even later), details will be forgotten and intervention will not be possible. Study monitoring questionnaires for the ISSP (completed well after question design and translation have been completed) sometimes contain documentation on translation challenges for two or three phrases. The templates used in recent German ISSP translation discussions note a myriad of challenges [2].

- Any changes countries make to their design protocols and procedures and any reservations they have about development must be carefully documented. If these are made available in a timely fashion to either the questionnaire drafting coordinator or, as appropriate, the central coordinating center, problems can be addressed. For example, feedback to drafting groups from countries participating in the ISSP and ESS studies sometimes lead to changes in draft versions of source questions.

- At a later stage, documentation might be helpful in understanding potential differences in the data, either over the course of the study (within a country) or across variables (between countries).

- Providing tools to make the job easier encourages people to engage in the task and ensures better documentation.

- Demonstrating the importance of documentation motivates people to engage in it. Even simple things can help convince and motivate ─ for example, showing how a template can help check for flipped order of answer categories across a range of questions.

Appendix A

Some advantages and constraints on different approaches to question design

Approach	Advantages	Constraints
Ask the same question and translate (ASQT)	If successful, questions and item scales can be compared one-by-one across data sets and thus permit the most sophisticated analyses based on what is sometimes called full scalar equivalence (see [25]).	Developing for an ASQT questionnaire may result in reduced specificity of questions used and a resultant loss of saliency and fine-grained information.
	ASQT is the least complicated approach to organize and implement. This is not to suggest it does not involve considerable effort as reflected in this chapter and Translation.	Conceptual coverage for all or some of the populations in the study may thus be reduced and not comparable across populations.
	Researchers engaged in it and clients requesting or expecting it may feel more like they are on familiar territory with this model than with others.	At the translation stage, those implementing the ASQT may have inappropriate goals for the translation product and produce poor target language versions.
	ASQT potentially permits replication of existing questions—provided their basic suitability for translation and fielding with the target populations is ensured.	Replicated questions encourage close translation and may not be optimal for one or more target populations.
		ASQT does not work well for some kinds of questions at all (e.g., background variables such as education).

		ASQT does not work well for development of comparable answer scales across languages.
		ASQT or ASQ and adapt approaches call for expertise in question development and translation in areas still requiring basic research and/or training.
Decentering	Allows two questionnaires to be developed in collaboration and creates the potential for full scalar equivalence (see [25]).	May result in questions with low saliency for either culture since anything that constitutes a problem in the course of development is removed or altered. This would be an obstacle to full scalar equivalence.
	Avoids the situation where the needs of one questionnaire and language/culture dominate.	Decentering is not viable for multilanguage projects.
	Can be useful in developing comparable questions for very disparate cultures.	Decentering is very work intensive and there is little information about recent experiences using this technique.
Ask different questions approaches	Avoids the need to base questionnaires for various cultures and languages on translation of a source questionnaire.	Little detailed information is available about recent projects adopting an ADQ approach. Researchers have few guidelines about how to develop the quality assurance and control steps needed.

	Researchers can select the indicators and questions considered most salient for a given population provided these produce data which can still be compared. across populations.	If different populations are only asked questions developed for them, item-by-item analyses across populations are more difficult to justify.
	It is easier for a group joining an ADQ-based study after other groups have developed and fielded their questionnaires to produce a suitable questionnaire for their context than it is for researchers joining and ASQT project after the source questionnaire has been finalized.	Most researchers and clients are unfamiliar with ADQ approaches.
Mixed approaches combining ASQ and ADQ components	These can combine the advantages of ASQ and ADQ.	They increase the number and kind of procedural steps to be implemented and assessed.
		They call for expertise in areas still requiring basic methodological research.

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Ask different questions (ADQ)	An approach to question design where researchers collect data across populations or countries based on using the most salient population-specific questions on a given construct/research topic. The questions and indicators used in each location are assumed (or better, have been shown) to tap a construct that is germane or shared across populations.
Ask the same questions (ASQ)	An approach to question design whereby researchers collect data across populations/countries by asking a shared set of questions. The most common way to do this is by developing a source questionnaire in one language and then producing whatever other language versions are needed on the basis of translation or translation and adaptation. Hence the description used in the chapter of "ASQ and translate". Decentering is a second way to "ask the same questions" but this procedure is differently organized.
Attitudinal question	A question asking about respondents' opinions, judgments, emotions, and perceptions. These cannot be measured by other means; we are dependent on respondents' answers. Example: *Do you think smoking cigarettes is bad for the smoker's health?*
Behavioral question	A question asking respondents to report behaviors or actions. Example: *Have you ever smoked cigarettes?*
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.

Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: *Do you smoke?* *Yes* ___ *No* ___
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Cognitive interview	A pretesting method designed to uncover problems in survey items by having respondents think out loud while answering a question or retrospectively.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Decentering	An approach to designing questions in two languages in which neither the language nor culture involved is allowed to dominate. A Ping-Pong-like process of formulation and comparison between the two languages is used to develop versions in each language. Any language or cultural obstacles met with are resolved, often by removing or changing wording in one or both languages. The question formulation in both languages then moves on from that modification. Since the process removes culture-specific elements from both versions, decentered questions may be vague and not especially salient for either target population.

Differential item functioning (dif)	Item bias as a result of systematic differences in responses across cultures due to features of the item or measure itself, such as poor translation or ambiguous wording.
Embedded experiments	Embedded experiments are included within the framework of an actual study.
Factual question	A question that aims to collect information about things for which there is a correct answer. In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative records. Factual questions can be about a variety of things, such as figure-based facts (date, age, weight), events (pregnancy, marriage), and behaviors (smoking or media consumption). Example: *Do you smoke?*
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Focus group	Small group discussions under the guidance of a moderator, often used in qualitative research that can also be used to test survey questionnaires and survey protocols.
Interpenetrated sample assignment, interpenetration	Randomized assignment of interviewers to subsamples of respondents in order to measure correlated response variance, arising from the fact that response errors of persons interviewed by the same interviewer may be correlated. Interpenetration allows researchers to disentangle the effects interviewers have on respondents from the true differences between respondents.
Item Response Theory (IRT)	A theory that guides statistical techniques used to detect survey or test questions that have item bias or differential response functioning (see dif). IRT is based on the idea that the probability of a response an individual provides is a function of the person's traits and characteristics of the item.

Language harmonization	Language harmonization can be understood as the procedures and result of trying to find a common version (vocabulary and/or structure) across questionnaires for different regional varieties of a "shared" language.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Mode	Method of data collection.
Multi-Trait-Multi-Method (MTMM)	A technique that uses the correlations between multiple methods (i.e. modes) and multiple traits (i.e. variables) to assess the validity of a measurement process.
Open-ended question	A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers. Example: *What is your occupation?* *Please write in the name or title of your occupation*___________
Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.

Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Ranking format	A response format where respondents express their preferences by ordering persons, brands, etc. from top to bottom, i.e., generating a rank order of a list of items or entities. Example: *Listed below are possible disadvantages related to smoking cigarettes. Please enter the number 1, 2, 3, or 4 alongside each possible disadvantage to indicate your rank ordering of these. 1 stands for the greatest disadvantage, 4 for the least disadvantage.* _____ *Harmful effects on other people's health* _____ *Stale smoke smell in clothes and furnishings* _____ *Expense of buying cigarettes* _____ *Harmful effects on smoker's health*

Rating format A response format requiring the respondent to select one position on an ordered scale of answer options.
Example: *To what extent do you agree or disagree with the following statement?*

It is a good idea to ban smoking in public places.
Strongly agree
Somewhat agree
Neither agree nor disagree
Somewhat disagree
Strongly disagree

Reliability The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.

Replicated question A question which is repeated (replicated) at a later stage in a study or in a different study. Replication assumes identical question wording. Questions which were used in one study, then translated and used in another are also frequently spoken of as having been "replicated."

Response styles Consistent and stable tendencies in response behavior which are not explainable by question content or presentation. These are considered to be a source of biased reporting.

Sample design Information on the target and final sample sizes, strata definitions and the sample selection methodology.

Sample element A selected unit of the target population that may be eligible or ineligible.

Sampling frame A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.

Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Socio-demographic question	A question that typically asking about respondent characteristics such as age, marital status, income, employment status, and education. Example: *What year and month were you born?*
Source language	The language in which a questionnaire is available from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Source questionnaire	The questionnaire taken as the text for translation.
Stand-alone experiment	An experiment conducted as an independent research project.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.

Tailoring	The practice of adapting interviewer behavior to the respondent's expressed concerns and other cues, in order to provide feedback to the respondent that addresses his or her perceived reasons for not wanting to participate.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Validity	The extent to which a variable measures what it intends to measure.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Working group	Experts working together to oversee the implementation of a particular aspect of the survey lifecycle (e.g., sampling, questionnaire design, training, quality control, etc.)

References

[1] Baumgartner, H., & Steenkamp, J. B. (2001). Response style in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2)*,* 143–156.

[2] Behr, D. (2010). Translationswissenschaft und international vergleichende Umfrageforschung: Qualitätssicherung bei Fragebogenübersetzungen als Gegenstand einer Prozessanalyse [Translation research and cross-national survey research: quality assurance in questionnaire translation from the perspective of translation process research]. Mannheim, Germany: GESIS.

[3] Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.

[4] Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design—For market research, political polls, and social and health questionnaires*. San Francisco, CA: Jossey-Bass.

[5] Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage Publications.

[6] de Leeuw, E.D. (2008). Choosing the method of data collection. In E. D. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 113–135). New York/London: Lawrence Erlbaum Associates.

[7] Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York, NY: John Wiley & Sons.

[8] Fowler, F. J. Jr. (1995). *Improving survey questions: Design and evaluation* (Vol. 38, Applied social research methods series). Thousand Oaks, CA: Sage Publications.

[9] Granda, P., Wolf, C., & Hadorn, R. (2010). Harmonizing survey data. In J. A. Harkness, M. Braun, B. Edwards, T. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E., Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 315-334). Hoboken, NJ: John Wiley & Sons.

[10] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[11] Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. D. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 56–77)*.* New

York/London: Lawrence Erlbaum Associates.

[12] Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R., & Villar, A. (2010). Designing questionnaires for multipopulation research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E., Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural, and multiregional contexts* (pp. 33–58). Hoboken, NJ: John Wiley & Sons.

[13] Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 117-140). Hoboken, NJ: John Wiley & Sons.

[14] Harkness, J. A., van de Vijver, F. J. R., & Johnson T. P. (2003). Questionnaire design in comparative research. In J. A. Harkness, F. J. R. van de Vijver & P. M. Mohler (Eds.), *Cross-cultural survey methods* (pp. 19-34). Hoboken, NJ: John Wiley & Sons.

[15] Hoffmeyer-Zlotnik, J. H. P., & Wolf, C. (Eds.). (2003). *Advances in cross-national comparison: A European working book for demographic and socio-economic variables*. New York, NY: Kluwer Academic/Plenum Publishers.

[16] International Organization for Standardization. (2006). *Market, opinion and social research—terms, definitions and service requirements* (ISO 20252). Geneva. Retrieved April 30, 2010, from https://www.astandis.at/shopV5/Preview.action%3Bjsessionid=0D8F503B3BD48375403D96257760FABB?preview=&dokkey=342276&selectedLocale=en/

[17] Johnson, T. P. & van de Vijver, F. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. van de Vijver & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp.195–204). Hoboken, NJ: John Wiley & Sons.

[18] Kolsrud, K., & Kalgraff Skjåk, K. (2005). Harmonising background variables in the European Social Survey. In J. H. P. Hoffmeyer-Zlotnik & J. A. Harkness (Eds.), *Methodological aspects in cross-national research* (ZUMA-Nachrichten Spezial 11). Mannheim, Germany: ZUMA. Retrieved April 28, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial11.pdf

[19] Lyberg, L. E., & Stukel, D. M. (2010). Quality assurance and quality control in cross-national comparative studies. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in*

multinational, multicultural, and multiregional contexts (pp. 227–250). Hoboken, NJ: John Wiley & Sons.

[20] Mohler, P. Ph. (2006). Sampling from a universe of items and the de-machiavellization of questionnaire design. In M. Braun & P.Ph. Mohler (Eds.), *Beyond the horizon of measurement - Festschrift in honor of Ingwer Borg* (ZUMA-Nachrichten Spezial,10). Mannheim, Germany: ZUMA. Retrieved April 28, 2010, from http://www.gesis.org/fileadmin/upload/ forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial10.pdf

[21] Payne, S. L. (1980). *The art of asking questions.* Princeton, NJ: Princeton University Press.

[22] Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation.* Mahwah, NJ: Lawrence Erlbaum Associates.

[23] Schwarz, N, Oyserman, D., & Petycheva, E. (2010). Cognition, communication, and culture: Implications for the survey response process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural, and multiregional contexts* (pp. 177–190). Hoboken, NJ: John Wiley & Sons.

[24] Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.

[25] van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.

[26] Van Deth, J. W. (1998). *Comparative politics: The problem of equivalence.* London: Routledge.

[27] Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 398–420). New York, NY: American Museum of Natural History.

[28] Willimack, D.K., Lyberg, L. E., Martin, J., Japec, L., & Whitridge, P. (2004). Evolution and adaptation of questionnaire development, evaluation, and testing methods for establishment surveys. In S. Presser, J. M. Rothger, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 385–408). Hoboken, NJ: John Wiley & Sons.

[29] Yang, Y., Harkness, J. A., Ching, T-Y., & Villar, A. (2010). Response styles and culture. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith

(Eds.), *Survey methods in multinational, multicultural, and multiregional contexts* (pp. 203–226). Hoboken, NJ: John Wiley & Sons.

For information surveys provide about themselves, see the links below:

Afrobarometer Survey. Retrieved July 1, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

Asian Barometer. Retrieved July 1, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

European Social Survey. Retrieved July 1, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=23&Itemid=318

International Social Survey Programme. Retrieved July 1, 2010, from http://www.issp.org/

Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

Survey of Health, Ageing and Retirement in Europe. Retrieved July 1, 2010, from http://www.share-project.org

VII. Adaptation

Janet Harkness

Introduction

The term adaptation, as used in this chapter, refers to the deliberate modification of a question or questionnaire to create a new question or questionnaire.

At the same time, adaptation needs are considered at different times in different projects and it is likely that some adaptation needs only become apparent during translation or during pretesting of a translated questionnaire. It is therefore not possible, in terms of the survey lifecycle, to identify a single unique stage as the stage at which adaptation needs might be recognized or addressed.

Why adapt questions?

Adaptation may be made to the content, format, response options, or visual presentation of any part of a question, questionnaire, or instrument. The purpose of adaptation is to better fit the needs of a new population, location, language, or mode, or any combination of these ([6] [7] [10]; see also Instrument Technical Design and Translation: Language Harmonization).

When developing new studies, researchers frequently modify questions that have been used in other studies and then use these modified versions. The motivations for such modifications may or may not be documented. Sometimes changes are related to adapting to meet new needs. However, sometimes changes are made simply because those making the changes consider them to result in a generally "better" version or instrument. This chapter focuses only on changes made to meet new needs as described above (adaptations).

In one language contexts, questions and questionnaires may be deliberately adapted for a variety of reasons. In longitudinal surveys, for example, wording might be updated to stay abreast with current usage; "wireless" could be replaced by "radio" [12], for example. Wording might also be changed to better reflect current social realities, such as adding the Internet as an information source in media usage questions [12], or adding "surfing the Web" as a category in questions about pastimes and hobbies. Changes might also be changed to accommodate a new population; modifying vocabulary, presentation, and instructions to suit a child population rather than an adult one, for example.

In cross-cultural and multilingual projects, adaptation is often related to the need to translate a questionnaire into another language in order to study new populations. In the following chapters, the terms "source language" and "target language" are used to distinguish between the language translated out of (the source language) and the language translated into (the target language).

In some projects, adaptations may already be anticipated in the source questionnaire, that is, the questionnaire on which other language versions are based and derived. Thus a source questionnaire question about pride in one's nationality, "How proud are you to be [nationality]*?"* anticipates a country-specific adaptation inside the square brackets, with each participating country entering the relevant nationality (e.g., Spanish, German, Chinese) in the slot indicated by the square brackets in their version of the questionnaire.

Socio-demographic questions often require adaptations to be made in different locations and languages (see Translation: Language Harmonization).

The need to make some adaptations might only become apparent in the course of translating the source questionnaire into a given target language. This could be because features of the target language itself make adaptation necessary or because a translated version of the source question, although possible, would not achieve the required measurement goals.

Answer scales provide examples of adaptations occasioned by features of the target language. Agreement scale response categories developed in English frequently have a middle category "neither agree nor disagree." In languages such as Hebrew and Swahili, this phrase cannot properly be translated by simply translating the words. The closest semantic option available to translate "disagree" in Hebrew, for example, corresponds to "no agree." In addition, the words "neither" and "nor" are the same as the target language element corresponding to "no." Thus "neither agree nor disagree," if translated element for element, would produce something like "no agree, no no agree;" this makes little sense in Hebrew [8]. The Hebrew phrase thus used in ISSP studies for the category "neither agree nor disagree" corresponds to "in the middle."

Frequently adaptations are motivated less by features of the target language than by the need to fit social, cultural, or other needs of the new linguistic group to be studied. Examples of adaptation not directly related to linguistic considerations abound. A recent international project proposed fielding the question, "Can you lift a two liter bottle of water or soda…," in multiple countries. The source question itself was not developed cross-culturally (see Questionnaire Design). Several locations (countries) noted that (a) the normal size of bottle in their context was 1.5 liter bottles, not 2 liter bottles, (b) that they were unsure whether the bottle referred to was intended to be glass or plastic (which would affect the lifting task), (c) that "soda" was not a salient generic concept in their locations, and (d) that the formulation in English which indicates that the bottle is not empty ("bottle of water or soda") needed to become "a full bottle of water" or "a bottle full of water" in their translations. However, there was some concern that these much more explicit renderings of "bottle of water" might alter respondent perceptions of the lifting task.

At times, as reflected in these examples, the needs of translation and those of adaptation are entangled. Thus, the appropriate or viable translation for a given context may also be a translation that includes adaptation of content, format, or some other questionnaire feature. Translations of an American question referring to being able to walk "several blocks" also needed to adapt the phrase "several blocks" for Great Britain and provide the distance for European locations in terms of yards or meters [7]. It is not always possible, therefore, to distinguish neatly between translation needs and the need to adapt other features of the question or questionnaire.

Common forms of adaptation

The categories identified below are based on distinctions found in [6], [7], and [10].

System-driven adaptation

Units of measurement differ across countries and may require adaptation (e.g., Imperial [yards, pounds] versus Metric [meters, kilos]; Fahrenheit versus Celsius). Adaptations will need to be considered for any mention in instruments of length, area, dry volume, liquid capacity, weight or mass, and also currency.

Adaptation to improve or guide comprehension

In preparing to use the question, "Can you run 100 yards?" in Vietnam, local researchers worried that the distance would not be clear to Vietnamese respondents and adapted it to, "Can you run 100 yards or the distance of three light poles?" to help respondents envision the distance intended [5] [7]. In this particular example, the distance mentioned in the source version is retained but also supplemented by a localized indication of the intended distance.

Adaptation to improve conceptual coverage

Sometimes question components are added for a given location to better tap the intended dimension or construct. For example, the symptoms shown by patients with a given disease (as well as the treatments, the attributed causes, and the places to get help) can differ across cultures. Including mention of local symptoms, as relevant, can improve the accuracy of information collected at the local level and for the combined data set.

Adaptation related to cultural discourse norms

Speech communities differ in the way in which they frame and carry out communication. Depending on the culture and language involved, indicators of politeness or deference may be required in the interview script or the self-completion questionnaire (polite imperatives, acknowledgment of relative status of interviewer and respondent, apologies for asking a question, etc.).

In some contexts adaptations are made without the scientific community currently acknowledging these as part of questionnaire adaptation needs. For example, Korean is a language with a systematic honorifics system reflecting social status, age, interpersonal relationships between participants in a discourse, and, indeed, much more [13]. In interviewer-assisted applications, such discourse and etiquette requirements affect what interviewers say, depending on whom they are interviewing. In some diglossic linguistic contexts, the gap between written forms of a language and spoken forms can be quite large. This can mean that interviewers have a written script that conforms to the norms of the written standard of the language but are required**,** in "speaking the script," to conform to spoken norms of the language (see [9] and contributions in [11]).

Adaptation and cultural sensibilities

Cultural sensibilities with regard to a wide range of topics differ from culture to culture. Such differences motivated adaptations for Japan in the Picture Completion section of the Wechsler Intelligence Scale for Children (WISC-III). Instead of a depiction of a person with a body part missing, the Japanese version used an inanimate object with a part of that object missing [15].

Adapting design components or characteristics

Changes to the technical design of an instrument can be motivated by many factors. The direction languages are read or written in, a population's familiarity with certain visual representations (thermometers, ladders, Kunin faces), and a wide range of culturally anchored conventions related to visual presentation, including color symbolism, representational preferences, and conventions of emphasis, may call for adaptation of components of the source questionnaire (see also Instrument Technical Design).

Adaptation related to lexicon and grammar

The lexicon (a language's vocabulary) and grammar of a language may also make changes in design necessary. An example already discussed is the response category "neither agree nor disagree" which has been rendered in Hebrew International Social Survey Programme questionnaires as "in the middle."

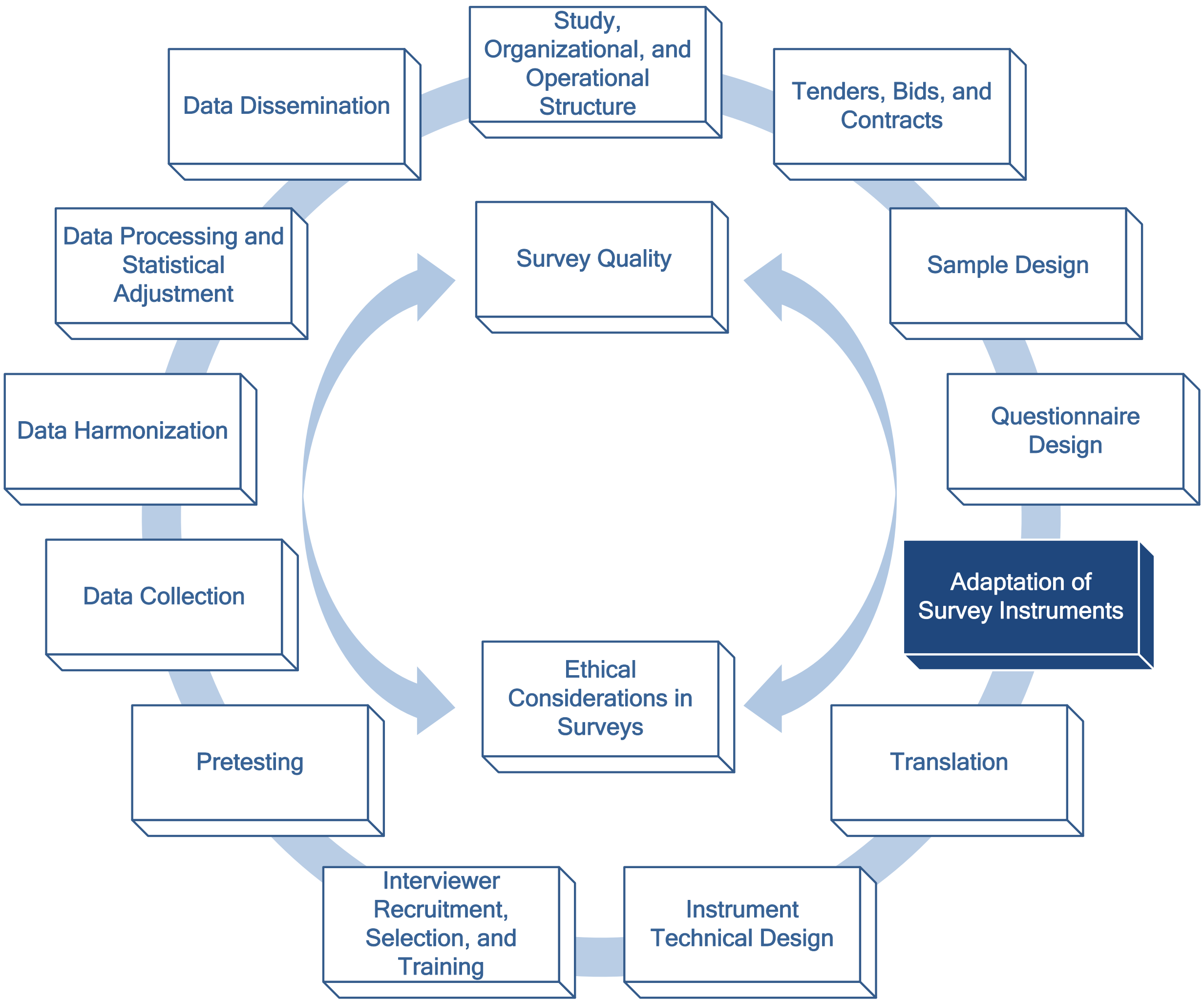
Adaptation to maintain or to reduce level of difficulty

Educational and cognitive ability tests are biased if it is easier for one population to answer correctly or perform a task required than it is for another population of equal ability on that item [16]. A wide range of question types is thus sometimes adapted to maintain the same level of difficulty across different populations. Research in educational and psychological testing discusses such issues (see, for example, [3] and [4]).

In studies of opinions, behaviors and attitudes, the goal is generally more one of keeping respondent burden low. Adjustments may thus sometimes be made to simplify the vocabulary used in a translation for populations with expected low levels of education or to increase instructions and explanations for those unfamiliar with the procedures of survey research. Response scale presentation is sometimes supplemented for populations unfamiliar with the notions of rating, for example, or for those unfamiliar with conceptualizing the response scale concepts in relation to entities asked about [2] [14].

Figure 1 shows the adaptation of survey instruments within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To make a survey instrument better fit the needs of a new population, location, language, or mode.

1. **Determine the policy, people, and procedures for adaptation for the project.**

Rationale

Adaptation needs will arise in most comparative projects and should be addressed professionally. Any quality assurance and quality monitoring framework must therefore include a plan for how to deal with adaptation. This plan should propose procedures to identify and address adaptation needs for each location and how to make decisions about documentation. It should also determine how any effort to coordinate adaptations or their documentation is to be organized (see [6]).

Procedural steps

- Plan coordination of adaptation development and the tools to be used to develop and document the process and outputs.

- Identify a suitable team with the necessary skills to work on adaptation problems (see Guideline 2 below).

- Decide on a procedure for approving adaptation by the persons assigned to decide and approve adaptations. In projects aiming to ask the same questions (ASQ) of each population, substantive adaptations should only be made if they are required to ensure comparable measurement or avoid some other important negative consequence.

- Determine the levels at which adaptations will be accommodated (only at target question level or also, as recommended here, at source question level (see Lessons Learned below).

- Decide on a strategy to ensure that participating groups (locations, countries, etc.) are informed about adaptations being proposed by other members and can contribute their own proposals or reactions.

Lessons learned

- By anticipating certain adaptations in an ASQ source and translate model, the translated version are likely to be more consistent with the measurement intended in the source questionnaire. However, it is very likely that some adaptation needs will not be recognized until translated versions are available.

2. Recruit a team to work on adaptations.

Rationale

Adaptations are made to address modifications necessary to be able to interview multiple populations. The spread of skills and range of cultural experience required cannot be provided by one person.

The team should bring together knowledge about and an understanding of (1) adaptation needs in general, (2) the types of adaptation that are usually needed, (3) the strategies commonly used to adapt, (4) measurement comparability needs, (5) language proficiency in whatever languages are involved, and (6) relevant cultural information.

Procedural steps

- Identify a small group of people who can, as a team, provide the skills and competencies needed for the six points mentioned above.
- Identify at least two people for each given location or instrument to supplement the team. These additional team members contribute only to the specific instrument they can provide input on. They provide the specific cultural awareness and language competence needed for a given location and language. However, issues identified for one location and population may prove relevant for others.
- Brief all team members on the goals of the adaptation steps, the procedures, any tools to be used, and the documentation required.

Lessons learned

- Briefing and providing examples of what is desired and not desired is important. Many members of such teams might be working consciously on adaptation for the first time. In addition, some team members with experience with adaptation might have learned practices the current team does not want to endorse. Providing examples for discussion during briefing and training reduces the likelihood of members making incorrect assumptions about what is required and how to proceed.

3. Review, as relevant, the source questionnaire for adaptation needs.

Rationale

Identifying adaptation needs in the source questionnaire may result in a better source questionnaire (that is, one that is easier to work with as a source questionnaire). By identifying elements to consider for adaptation in the source document, comparability across different questionnaire versions can also be enhanced.

Procedural steps

- Assign the work to a person or persons familiar with the common forms of adaptation in surveys, knowledgeable about the questionnaire as well as the measurement goals of each question, and with a good understanding of the cultural and social realities of both source and target populations. Provide a format for indicating potential adaptation elements.
- Keep a record of all elements identified and the rationale for these.
- Provide examples of what is required in terms of adaptation in the record.
- Check the suggestions made with a range of locations participating in the project. The members engaged for local consultation would be useful contacts for this.
- Adjust the adaptation proposals for the source questionnaire as seems appropriate.

Lessons learned

- It may not be easy to find people with experience in adaptation procedures. People with extensive experience in drafting questionnaires for multicultural projects and translators may be good first choices; each can provide different insights based on their different knowledge and experience.
- The ability to look at a questionnaire with an awareness of other cultures' needs can be trained but it needs to be based on some background of cross-cultural experiences and awareness. Translators develop the ability to think across and between cultures in the course of their training. Their insights and their explication of motivations for suggested changes could help others in the team learn what is needed. At the same time, translators cannot be expected to understand all the measurement factors to be considered in question adaptation. In addition, translators are not necessarily in touch with the on-the-street reality of interviewing and the everyday language of the target population. This is why a team providing a spread of expertise is recommended.

4. Review the translated questionnaire or instrument for adaptation needs.

Rationale

A review with respect to adaptation can be incorporated into the translation phases. Some adaptation proposals are likely to result from the translation process in any case. However, some adaptation needs that are unrelated to translation may not be apparent to the translation team. It is, therefore, important to check for other adaptation needs once the translation is completed. In addition, the adaptation team may have access to knowledge about adaptation undertaken in other languages involved in the project that an individual translation team does not.

Procedural steps

- Engage at least two people for a given location and language to work with the adaptation team. The persons chosen should, together, provide language and translation skills and a good understanding of the cultural contexts of target versions. The core team producing the local target version of the questionnaire could help them as necessary to be aware of source version implications and cultural assumptions inherent in it. These people need not be extremely proficient in the language of the source questionnaire. If suitable local people are readily available, using two different people from those advising on adaptation for the source questionnaire (Guideline 3.1) could minimize repetition and transfer of topics from the source questionnaire review to the current review.

- Provide a format for indicating potential adaptation elements, along with examples.

- Keep a record of all elements identified and the rationale for these.

- Check the suggestions made by the adaptation team with groups formed from other locations and adjust the adaptation proposals accordingly. This step might best be undertaken as a late step in deciding adaptations for the entire project.

Lessons learned

- Given the meager literature on the rationale and procedures of adaptation in surveys, adaptation teams may end up making decisions based on common sense and best guesses. Pretesting adaptation decisions before implementation is thus essential.

5. **Document adaptations and the rationale for making them.**

Rationale

Documentation of adaptations is important for version control across locations and adaptations in one round of a survey. It also makes it possible to check content and presentation through any longitudinal iterations of a survey or a question. Such documentation can also ultimately inform the development of a more refined understanding of adaptation practices.

Lessons learned

- Ensure the documentation of changes and their rationale is made publicly available. At the moment it is not easy to find literature on adaptation that presents procedures and motivations in detail. The documentation taken by teams as proposed above will form an important basis for surveys in the future and help advance this area of methodology.

- The motivation for adaptations may also not be evident to those not involved in the adaptation process. Secondary analysts, for example, would benefit from a record of the rationale behind adaptations.

6. **Test adaptations made with the target population.**

Rationale

Adaptation results in new questions. New questions should be tested with people representative of the target population.

Procedural steps

- Pretest adapted instruments to find out whether the questions are understood as intended and can be answered without undue burden.

- Include quantitative assessment (see Pretesting).

Lessons learned

- It is important to streamline development of adapted instruments as much as possible in order to have enough time and resources to undertake the various steps and testing of these steps. Adaptation needs should be considered at each stage of development. Development and pretesting of the source questionnaire should keep adaptation needs in mind. The question about being able to lift a 2 liter bottle of water or soda, for example, could have been evaluated in terms of the availability of bottled beverages, the saliency of the size of the bottles, and the material of which they might be made. Translation alone cannot remedy such matters.

- If adaptation is left until the last moment, there may be no more time or resources to pretest.
- If sharing findings and conclusions about adaptation across locations involved in a project is not organized in an efficient and timely fashion, individual locations are not able to benefit from solutions or problems found in other locations.
- Extensive evaluations of various kinds are needed to establish whether adapted or translated questions result in comparable measurement. The health-related quality of life literature on translated instruments, even on just the SF-36 Health Survey, is revealing in this respect. See, for example, [1] and references cited there.

Accuracy	The degree of closeness an estimate has to the true value.
Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Ask the same questions (ASQ)	An approach to question design whereby researchers collect data across populations/countries by asking a shared set of questions. The most common way to do this is by developing a source questionnaire in one language and then producing whatever other language versions are needed on the basis of translation or translation and adaptation. Hence the description used in the chapter of "ASQ and translate". Decentering is a second way to “ask the same questions” but this procedure is differently organized.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Contract	Defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.

Decentering	An approach to designing questions in two languages in which neither the language nor culture involved is allowed to dominate. A Ping-Pong-like process of formulation and comparison between the two languages is used to develop versions in each language. Any language or cultural obstacles met with are resolved, often by removing or changing wording in one or both languages. The question formulation in both languages then moves on from that modification. Since the process removes culture-specific elements from both versions, decentered questions may be vague and not especially salient for either target population.
Diglossic linguistic contexts	Diglossic linguistic contexts exist in communities that use two or more markedly different varieties of a language in different contexts. The variety used may be determined by whether the language is written or spoken in a given instance or by the relationships between participants in a discourse. Considerations such as age, gender, social status, and the topic under discussion may all contribute to the form chosen in any given instance.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Longitudinal study	A study where elements are repeatedly measured over time.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Mode	Method of data collection.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.

Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Questionnaire adaptation	The deliberate technical or substantive modification of some feature of a question, answer options, or other part of a questionnaire to better fit a new socio-cultural context or particular target population (updating language: "radio" for "wireless", adapting an adult questionnaire for children: "tummy" for "stomach"; or tailoring for cultural needs: walk several blocks versus walk 100 yards).

Rating format	A response format requiring the respondent to select one position on an ordered scale of answer options. Example: *To what extent do you agree or disagree with the following statement?* *It is a good idea to ban smoking in public places.* *Strongly agree* *Somewhat agree* *Neither agree nor disagree* *Somewhat disagree* *Strongly disagree*
Response options	The category, wording, and order of options given with the survey question.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.

Socio-demographic question	A question that typically asking about respondent characteristics such as age, marital status, income, employment status, and education. Example: *What year and month were you born?*
Source document	The original document from which other (target) documents are translated or adapted as necessary.
Source language	The language in which a questionnaire is available from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Source questionnaire	The questionnaire taken as the text for translation.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.

References

[1] Bolton, P., & Tang, A. M. (2002). An alternative approach to cross-cultural function assessment. *Social Psychiatry and Psychiatric Epidemiology*, *37*(11), 537–543.

[2] Bullinger, M., Kirchberger, I., & Ware, J. (1995). Der deutsche SF-36 Health Survey (Űbersetzung und psychometrische Testung eines krankheitstibegreifenden Instruments zur Erfassung der gesundheitsbezogenen Lebensqualität [The German SF-36 Health Survey: Translation and psychometric testing of a generic instrument for the assessment of health-related quality of life]. *Zeitschrift für Gesundheitswissenschaften*, *3*(1), 21–36.

[3] Georgas, J., Weiss, L. G., van de Vijver, F. J. R. &. Saklofske, D. H. (Eds.). (2003). *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 215–225). Amsterdam: Academic Press.

[4] Hambleton, R. K., Merenda, P. F. & Spielberger, C. D. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ.: Lawrence Erlbaum Associates.

[5] Hanh, V. T. X., Guillemin, F., Cong, D. D., Parkerson, G. R. Jr., Thu, P. B., Quynh, P. T., & Briancon, S. (2005). Health related quality of life of adolescents in Vietnam: Cross-cultural adaptation and validation of the Adolescent Duke Health Profile. *Journal of Adolescence, 28*(1)*,* 127–146.

[6] Harkness, J.A. (in progress). *Towards guidelines on survey adaptation*. Paper for presentation at the XVII ISA World Congress of Sociology, Gothenburg, Sweden, July, 2010.

[7] Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. D. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 56–77)*.* New York/London: Lawrence Erlbaum Associates.

[8] Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. van de Vijver & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: John Wiley & Sons.

[9] Harkness, J. A., Schoebi, N., Joye, D., Mohler, P., Faass, T., & Behr, D. (2008). Oral translation in telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas et al. (Eds.), *Advances in telephone survey methodology* (pp. 231–249). Hoboken, NJ: John Wiley & Sons.

[10] Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P.

Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 117-140). Hoboken, NJ: John Wiley & Sons.

[11] Paulston, C.B. & Tucker, G.R. (Eds.). (2003). *Sociolinguistics: The essential readings*. Oxford: Blackwell.

[12] Smith, T. W. (2005). The laws of studying societal change. *Survey Research, 36*(2), 1-5. Retrieved April 28, 2020, from http://www.srl.uic.edu/Publist/Newsletter/2005/05v36n2.pdf

[13] Strauss, S. & Eun, J.O. (2005). Indexicality and honorific speech level choice in Korean. *Linguistics*, *43*(6), 611–651.

[14] Struwig, J., & Roberts, B. (2006, December). *Data collection methodology in South Africa*. Paper presented at the Sixth ZUMA Symposium of Cross-Cultural Survey Methodology, Mannheim, Germany.

[15] Ueno, K., & Nakatani, I. (2003). Japan. In J. Georgas, L. G. Weiss, F. J. R. van de Vijver & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 215–225). Amsterdam: Academic Press.

[16] van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.

VIII. Translation

Janet Harkness

Introduction

Following terminology used in the translation sciences, this chapter distinguishes between "source languages" used in "source questionnaires" and "target languages" used in "target questionnaires." The language translated out of is the source language; the language translated into is the target language.

Translation procedures play a central and important role in multilingual survey projects. Although good translation products do not assure the success of a survey, badly translated questionnaires can ensure that an otherwise sound project fails because the poor quality of translation prevents researchers from collecting comparable data.

The guidelines presented below envisage a team translation approach for survey instrument production. Evidence is growing that such procedures are effective for survey translation [4] [5] [9] [18] [20]. The guidelines address, at a general level, the steps and protocols recommended for survey translation efforts conducted using a team approach.

Before discussing team translation procedures, the chapter briefly outlines other approaches sometimes followed to produce or check survey translations and indicates why these are not recommended here. For discussion see [4] [5] [9] and [12].

Machine translation

Survey questions are a complex text type with multiple functions and components [3] [6] [11] [12]. As a result, any reduction of human involvement in the decision-making process of survey translation is ill-advised [12].

One of the main goals of machine translation, however, is to greatly reduce human involvement in translation production.

Do-it-yourself *ad hoc* translation

It is a mistake to think that because someone can speak and write two languages he or she will also be a good translator for these languages. Translation is a profession with training and qualifications. Translatology (under various names) is a discipline taught at the university level. Students of the translation sciences learn an array of skills and procedures and become versed in translation approaches and theories which they employ in their work. At the same time, as explained in the description of team translation following here, survey translation calls for not only a good understanding of translation but also of the business of survey measurement and how to write good questions. Under normal circumstances, a trained translator should not be expected to have a strong understanding of survey practices and needs, hence the need for a team of people with different skills [1] [5] [6] [7] [11] [12].

Unwritten translation

Sometimes bilingual interviewers translate for respondents as they conduct the interview. In other words, there is a written source questionnaire that the interviewers look at but there is never a written translation, only what they produce orally on the spot. This is sometimes called "on sight" translation, "on the fly translation," or "oral translation."

Another context in which survey translation is oral is when interpreters are used to mediate between an interviewer speaking language A and a respondent speaking language B. The interviewer reads aloud the interview script in language A and the interpreter is expected translate this into language B for the respondent. The interpreter is also expected to translate everything the respondent says in language B into language A for the interviewer. Research directly on the process of oral translation in surveys and how this affects interpretation, understanding, and data is quite sparse. Evidence available from recent investigations suggests that these modes of translation must be avoided whenever possible and that extensive training and briefing should take place if they must be used [10] [15] [16].

Translation and back translation

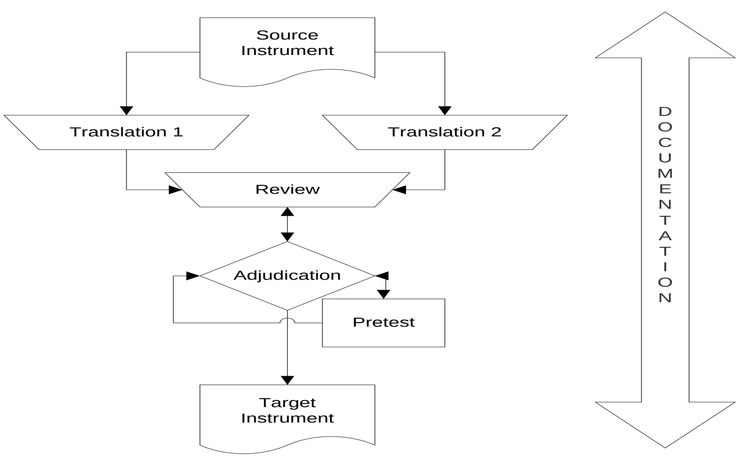
Even today, many projects rely on procedures variously called "back translation" to check that their survey translations are adequate. In its simplest form, this means that the translation which has been produced for a target language population is re-(or back-) translated into the source language. The two source language versions are compared to try to find out if there are problems in the target language text. As argued elsewhere, instead of looking at two source language texts, it is much better in practical and theoretical terms to focus attention on first producing the best possible translation and then directly evaluating the translation produced in the target language, rather than indirectly through a back translation. Comparisons of an original source text and a backtranslated source text provide only limited and potentially misleading insight into the quality of the target language text [7] [11] [12] [13] [14].

Introduction to team translation

In a team approach to survey translation, a group of people work together. Translators produce draft translation, reviewers review translations with the translators, one (or more) adjudicator decides whether the translation is ready to move to detailed pretesting and also decides when the translation can be considered to be finalized and ready for fielding.

Figure 1 below presents the TRAPD (Translation, Review, Adjudication, Pretesting, and Documentation) team translation model. In TRAPD, translators provide the draft materials for the first discussion and review with an expanded team. Pretesting is an integral part of the TRAPD translation development. Documentation of each step is used as a quality assurance and monitoring tool and each step of the translation effort includes assurance and monitoring elements [5] [6] [7] [12].

Figure 1. The TRAPD Team Translation Model



Procedures are partially iterative in team translation. The review stage reviews and refines draft translations. Adjudication, often a separate step from review, can lead to further modifications of the translation before it is signed off for pretesting (see Pretesting). Pretesting may again result in modifications before the adjudicator signs off on the version for final fielding.

Team approaches to survey translation and translation assessment have been found to be particularly useful in dealing with the fairly unique challenges of survey translation. The team can be thought of as a group with different talents and functions, bringing together the mix of skills and discipline expertise needed to produce an optimal version in the survey context where translation skill alone is not sufficient. Other approaches include having a single translator deliver a translation to the researchers or survey organization or having a translation agency deliver a translation to researchers. Such procedures are not specifically designed to bring together translators with other relevant experts in reviewing the translation in the way a team translation does. Further consideration of advantages that team efforts have over other approaches can be found in [3] [4] [5] [9] and [11].

Each stage of the team translation process builds on the foregoing steps and uses the documentation required for the previous step to inform the next. In addition, each phase of translation engages the appropriate personnel for that particular activity and provides them with relevant tools for the work at hand. These tools (e.g., documentation templates; see Appendix A) increase process efficiency and make it easier to monitor output. For example, translators producing the draft translations are required to keep notes about any queries they have on their translations or the source text. These notes are considered along with the translation output during the next review stage in which reviewers work together with the translators [5] [6] [7].

Team translation efforts work with more than one translator. Translators produce translation material and attend review meetings. Either each translator produces a draft version of the source questionnaire (double or full translation) or each translator gets parts of the source questionnaire to translate (split translation) [5] [11] [19]. The double translations or the sections of the split translation are refined in the review stage and possibly again after subsequent steps, as just described.

Whenever possible, translation efforts that follow a team approach work with more than one draft version of the translated text. A sharing of these draft versions and discussion of their merits is a central part of the review process. Two draft translations, for example, can dispel the idea of there only being one "good" or "right" translation. They also ensure that more than one translation is offered for consideration, thus enriching the review discussion. This encourages a balanced critique of versions [1] [5] [9] [17]. Contributions from more than one translator also make it easier to deal with regional variance, idiosyncratic interpretations, and translator oversight [5] [6] [11].

Survey translations also often call for sensitivity for words people speak rather than words people write. Apart from ensuring the needed spread of survey expertise and language expertise, the discussion that is part of team approaches is more likely to reveal vocabulary or vocabulary level/style (register) problems which might be overlooked in a review made without vocalization. Pretesting may, of course, reveal further respondent needs that “experts” missed.

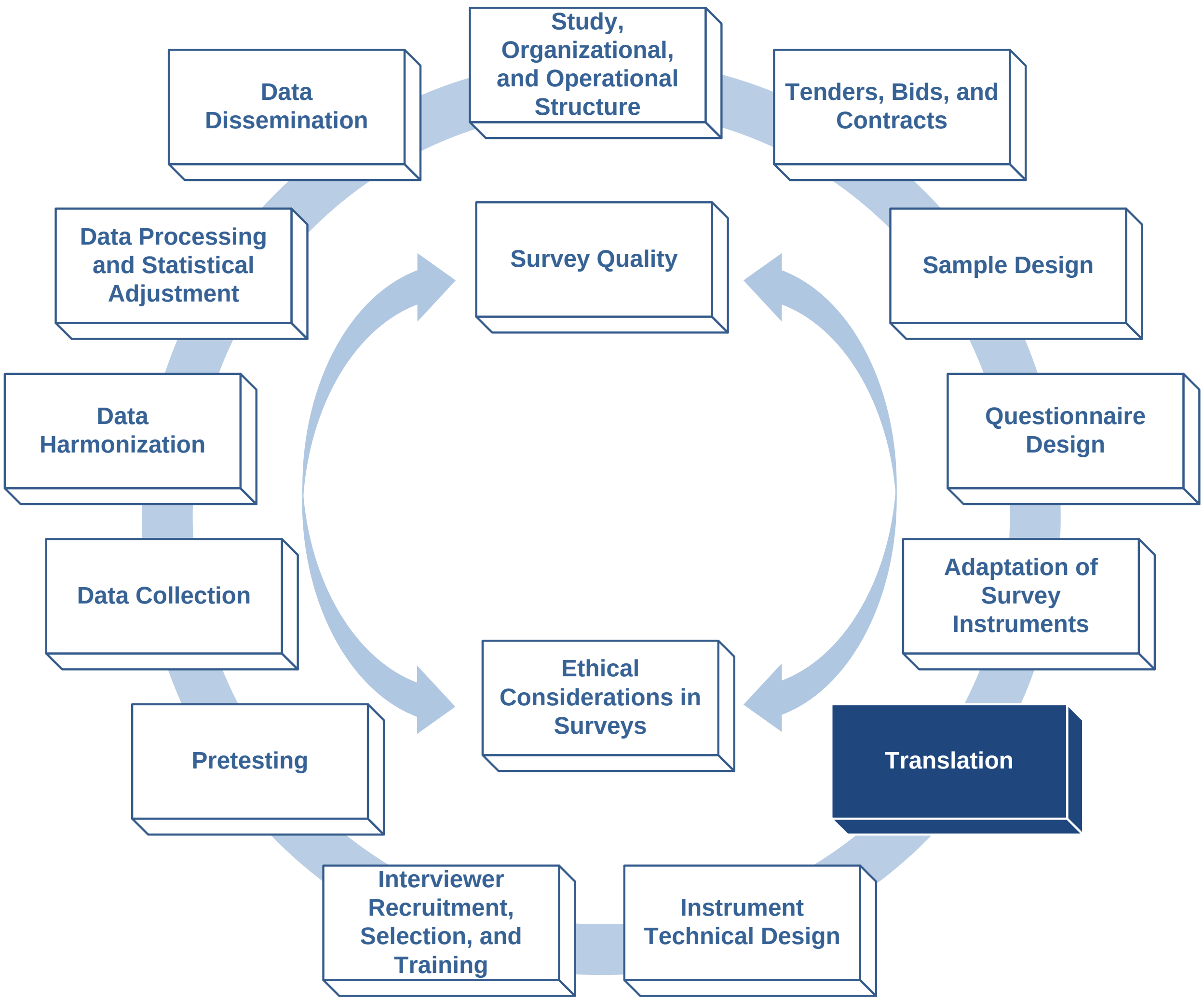
As noted, team-based approaches aim to include the translators in the review process. In this way, the additional cost of producing two draft translations would be offset by the considered contributions the translators can bring to review assessments. Since they are already familiar with the translation challenges in the texts, they make the review more effective. Split translation arrangements can still capitalize on the advantages of having more than one translator in the review discussion but avoid the cost of full or double translations. The advantages and disadvantages of each approach are discussed under guidelines 3 and 4 (see too, [5] and [19]).

The specifics of team translation procedures are considered below. For other aspects of translation production, please refer to:

Finding, Selecting, and Briefing Translation Team Members
Translation Management and Budgeting
Translation Tools
Translation Assessment
Language Harmonization
Translation Scheduling

Figure 2 shows translation within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 2. The Survey Lifecycle



Guidelines

Goal: To create and follow optimal procedures to standardize, assess, and document the processes and outcomes of survey questionnaire translation.

1. **Plan translation as an integral part of the study design.**
   This planning should include all the elements that will be part of the translation procedures (e.g., selection of team members, language harmonization) and should accommodate them in terms not only of procedural steps but with regard to hiring, training, budgeting, time schedules, and the questionnaire and translation production processes.

Rationale

Survey translation efforts are part of the target language instrument development and should be treated accordingly. In addition, when translations are produced in order to take part in a larger comparative project, forethought and a clear direction to planning and implementing translation will help produce translations across multiple locations which comply with project requirements.

.

Procedural steps

- Define the following:
  - The larger vision (e.g., a successfully implemented survey).
  - The concrete goal (e.g., a well-developed translation for the various contexts and populations).
  - Important quality goals (e.g., a population-appropriate translation, comparability with source questionnaire, efficiency and feasibility of translation procedures, timeliness).
  - Relevant factors (e.g., schedules, budget, personnel available, unexpected events).
  - Tasks involved (e.g., assembling personnel and the translation documents; preparing tools, such as templates; training personnel; producing and reviewing translations; pretesting; copyediting).

- Identify core team members (those people required for the team translation effort). See Appendix B for specific tasks of each core team member and other team players identified below.
  - Translators
  - Reviewer(s)
  - Adjudicator(s)
  - Copyeditor(s)

- Identify any other team players who may be required, based upon the size of the project, the mode of data collection, etc.
  - Co-coordinator
  - Substantive experts
  - Programmers
  - Other experts, such as visual design experts, adaptation experts
  - External assessors
  - Back-up personnel

- Determine whether regional variance in a language or shared languages need to be accommodated; decide on strategies for this as needed (see Language Harmonization).

- Select, brief, and train personnel (see Finding, Selecting, and Briefing

Translation Team Members). Identify the in-house and external staff and consultant needs on the project and follow appropriate selection, briefing, and training procedures for each person or group.

- Identify, acquire, and prepare the materials for translation.
  In addition to the source questionnaire, these may include advertising material, interviewer manuals, programmer instructions, and any supporting materials such as "showcards,", as well as statements of informed consent.

- Clarify payment arrangements for all involved (see Translation Management and Budgeting).

- Create a time schedule and identify project phases and milestones for members of the team (see Translation Management and Budgeting).

- Arrange for back-up team members in the event of unavailability or illness.

- Decide on the mode and schedule of meetings (face-to-face, web casting, or conference calls) and materials to be used at meetings (e.g., shared templates, software tools, documents deposited in e-room facilities, paper-and-pencil note-taking).

- Decide on other communication channels and lines of communication (reporting delays, illness, completion, deadlines).

- Decide whether each translator will prepare a full translation (double translation) or whether the material to be translated will be divided among the translators (split translation).

Lessons learned

- In major efforts, the bigger picture must first be considered to confirm which routine or special tasks are vital and which are not. It is easy to focus on procedures which are familiar and thus inadvertently miss other vital elements. For example, if consistency in terminology across versions is not something a project leader has usually considered, procedures to check for this might be overlooked in planning.

- The number of translations required varies among multilingual survey projects. The Afrobarometer Survey [21], the Asian Barometer Survey [22], and the European Social Survey Source [23] specify that every language group that is likely to constitute at least 5% of the sample should have a translated questionnaire.

- Planning and quality assurance and quality control should go hand-in-hand. When planning the project or procedure, it is also time to plan the quality assurance and quality control steps. For example, in planning the translation of answer scales, steps to check that scales are not reversed or a response category omitted can be incorporated into a translation template.

2. Have two or more translators produce first draft translations. If possible, have each translator produce a full translation; if that is not possible, aim to create overlap in the split translation sections each translator produces.

Rationale

Having more than one translator work on the draft translation(s) and be part of the review team encourages more discussion of alternatives in the review procedure. It also helps reduce idiosyncratic preferences or unintended regional preferences. In addition, including the translators in the review process who produced the initial drafts not only improves the review but may speed it up as well.

Procedural steps

- Determine lines of reporting and document delivery and receipts.
  - Translation coordinators typically deliver materials to translators. Coordinators should keep records of the delivery of materials and require receipt of delivery. This can be done in formal or less formal ways, as judged suitable for the project complexity and the nature of working relationships.
  - The project size and complexity and the organizational structure (whether centralized, for example) will determine whether translation coordinators or someone else actually delivers materials and how they are delivered.

- Determine the protocol and format for translators to use for note-taking and providing comments on source questions, on adaptations needed, and translation decisions. See Appendix A for documentation templates.

- Establish deadlines for deliveries, including partial translations (see below), and all materials for the review session.
  - Require each translator to deliver the first 10% of his/her work by a deadline to the coordinator (senior reviewer or other supervisor) for checking. Reviewing performance quickly enables the supervisor to modify instructions to translators in a timely fashion and enables hiring decisions to be revised if necessary.

  - Following the established protocol for production procedures and documentation, each translator produces his/her translation and delivers it to the relevant supervisor.

- Either have translators produce a full draft of the questionnaire and other materials to be translated or require each to produce some portion of the material (double or full translation or split translation).

- After receiving the translated materials, have the coordinator/senior reviewer prepare for the review session by identifying major issues or discrepancies in advance.

- Develop procedures for recording and checking consistency across the questionnaire at the finish of each stage of review or adjudication. (See Appendix A for documentation examples.)

Lessons learned

- The more complex the project (e.g., number of translations), the more careful planning, scheduling, and documentation should be (see Translation Management and Budgeting).

- Since the aim of review is to improve the translation wherever necessary, discussion and evaluation are at the heart of the review process. The senior reviewer or coordinator of the review meetings must, if necessary, help members focus on the goal of improvement. In line with this, people who do not respond well to criticism of their work are not likely to make good team players for a review.

- Review of the first 10% of the draft translation may indicate that a given translator is not suitable for the project because it is unlikely that serious deficiencies in translation quality can be remedied by more training or improved instructions. If this is the case, it is probably better to start over with a new translator.

- Draft translation is only the first step in a team approach. Experience shows that many translations proposed in drafts will be changed during review.

- If translators are new to team translation or the whole team is new, full translation rather than a split procedure is recommended whenever possible to better foster discussion at the review and avoid fixation on "existing" text rather than "possible" text.

- Translations that are fine from the translation point of view may not be strong enough from a measurement perspective. For instance, a translator might be inclined to reduce the numerous adverbial references in the following: "Generally speaking, how much television do you usually watch all in all on an average weekday?" Translators might feel that any one of “generally speaking,” “usually,” or “all in all,” or possibly the adjective “average” could usefully be omitted to make the sentence clearer. Rightly or wrongly, the question designer presumably felt it important to include each of these phrases to “guide" the respondent in what to consider. As things currently stand, there is little basic research into what and how respondents specifically process questions with such multiple "signposts."

- It is important to inform team members that changes to draft translations are the rule rather than the exception. The aim of a review is to review AND improve translations. Changes to draft translations should be expected and welcomed.

- Providing templates to facilitate note-taking will encourage team members to do just this. Notes collected in a common template can be displayed more readily for all to see at meetings.

- It may seem cheaper only to work with one translator and to eschew review sessions, since at face value, only one translator is paid for his or her translation and there are no review teams or team meetings to organize and budget for. In actuality, unless a project takes the considerable risk of just accepting the translation as delivered, one or more people will be engaged in some form of review.

- A professional review team may involve more people and costs than an ad hoc informal review but it is a central and deliberate part of quality assurance and monitoring in the team translation procedure. In addition, even in a team translation procedure, translation costs will make up a very small part of a survey budget and cannot reasonably be looked at as a place to cut costs. Experience gained in organizing translation projects and selecting strong translators and other experts is likely to streamline even these costs (see Translation Management and Budgeting).The improvements that team translations offer justify the additional translator(s) and experts employed.

- The burden of being the only person with language and translation expertise in a group of multiple other experts can be extreme. If more than one translator is involved in review, their contributions may be more confident and consistent and also be recognized as such.

- When translators simply “hand over” the finished assignment and are excluded from the review discussion, the project loses the chance to have translator input on the review and any discussion of alternatives. This seems an inappropriate place to exclude translator knowledge.

- Relying on one person to provide a questionnaire translation is particularly problematic if the review is also undertaken by individuals rather than a team.

- Even if only one translator can be hired, one or more persons with strong bilingual skills could be involved in the review process. (The number might be determined by the range of regional varieties of a language requiring consideration for the translation. Bilinguals might not be able to produce a useable translation but could probably provide input at the review after having gone through the translation ahead of the meeting.)

3. If possible, have new teams work with two or more full translations.

Rationale

Having new teams work with two or more full translations is the most thorough way to avoid the disadvantages of a single translation. It also provides a richer input for review sessions than the split translation procedure, reduces the likelihood of unintentional inconsistency, and constantly prompts new teams to consider alternatives to what is on paper.

Procedural steps

- Have several translators make independent full translations of the same questionnaire, following the steps previously described in Guideline 2.

- At the review meeting, have translators and a translation reviewer and anyone else needed at that session go through the entire questionnaire, question by question. In organizing materials for the review, depending on how material is shared for discussion, it may be useful to merge documents and notes in the template (see Appendix A).

Lessons learned

- The translation(s) required will determine whether more than two translators are required. Thus if the goal is to produce a questionnaire that is suitable for Spanish-speaking people from many different countries, it is wise to have translators with an understanding of each major regional variety of Spanish required. If, as a result, 4 or 5 translators are involved, full translation can become very costly and splitting the translation material is probably the more viable option.

- Translators usually enjoy not having to carry sole responsibility for a version once they have experienced team work.

4. **To save time and funds, have experienced teams produce split translations.**

Rationale

Split translations can save time, effort, and expense. This is especially true if a questionnaire is long or multiple regional variants of the target language need to be accommodated [5] [11] [19].

Procedural steps

- Divide the translation among translators in the alternating fashion used to deal cards in many card games.
  - This ensures that translators get a spread of the topics and possibly different levels of difficulty present in the instrument text.
  - This is especially useful for the review session—giving each translator material from each section avoids possible translator bias and maximizes translator input evenly across the material. For example, the Survey on Health, Ageing, and Retirement in Europe (SHARE) questionnaire has modules on financial topics, relationships, employment, health, and other topics [24].
  - By splitting the questionnaire (more or less) page for page, each translator is exposed to trying to translate a variety of topics and better able to contribute directly during review as a result.
  - Whenever possible, divide the questionnaire up in a way that allows for some overlap in the material each translator receives (see the first two Lessons Learned for this guideline).
  - Keep an exact record of which translator has received which parts of the source documents.

- Have each translator translate and deliver the parts he/she has been given for the review meeting.

- Use agreed formats or tools for translation delivery for the review session. For example, if a template is agreed upon, then different versions and comments can be entered in the template to make comparison easier during review. (See examples in Appendix A.)

- Develop a procedure to check for consistency across various parts of the translation.

- At the review meeting, have translators and the review team go through the entire questionnaire. When organizing materials for the review, depending on how material is shared for discussion, it may be useful to merge documents and notes (see Appendix A).

- Take steps to ensure that material or terms which recur across the questionnaire are translated consistently. For example, it is conceivable that two translators translate the same expression and come up with suitable but different translations. Source instrument references to a person's (paid) work might be rendered with "employment" by one translator, with "job" by another, and with "profession" by a third. In a given context, more than one of these could be acceptable. If this is the case, choose one term and use it consistently.

Lessons learned

- It is often necessary to split the material to address issues of time, budget, or language variety. Even observing the card-dealing division of the material ([5] [19]), there is often no direct overlap in split translations between the material the different translators translate. Translators are thus less familiar with the challenges of the material that they did not translate than the sections they translated. This can reduce the detail of input at the question-by-question review meeting. The senior reviewer must therefore take care to stimulate discussion of any section(s) where only one translation version is available.

- Budget and schedules permitting, it is useful to create some modest overlap in material translated. This allows the review team, including translators, to have an increased sense of whether there are large differences in translating approaches between translators or in their understanding of source text components at the draft production level.

- Giving people time to prepare the materials for the review meeting and making sure that they prepare is important for the meeting's success. Ad hoc suggestions and responses to translations are usually insufficient.

- Consistency checks can ensure that one translator's translation can be harmonized with another translator's possibly equally good but different rendering.

- In checking for consistency, it is important to remember this procedure must not be only mechanical (for example, using a find function in software). The source text may use one and the same term in different contexts with different meanings, while other language versions may need to choose different terms for different contexts. The opposite may also hold. Automatic harmonization based on “words” is thus not a viable procedure. For example, the English word "government" may need to be translated with different words in another language depending on what is meant. In reverse fashion, English may use different words for different notions which are covered by a single word or phrase in other languages. Examples: English "ready" and "prepared" can in some circumstances be one word in German; "he" and "she" are differentiated in English but not in Turkish or Chinese.

- Checks for general tone consistency are also needed. There is, for instance, a difference in tone in English between talking about a person's "job" and a person's "profession," or in referring to a young person as a "child" or a "kid."

5. **Review and refine draft translations in a team meeting. Review meetings may be in person, virtual, or a mix of the two. The time involved depends upon the length and complexity of a questionnaire, the familiarity of the group with procedures, and disciplined discussion. The work may call for more than one meeting.**

Rationale

The team meeting brings together all those with the necessary expertise to discuss alternatives and collaborate in refining the draft translations—translation reviewers, survey experts, and any others that a specific project requires.

Procedural steps

- Make all the translated draft materials available to team members in advance of the review meeting(s) to allow preparation.

- Provide clear instructions to members on expected preparation for the meeting and their roles and presence at the meeting.

- Arrange for a format for translations and documentation that allows easy comparison of versions.

- Appoint a senior reviewer with specified responsibilities.

- Have the senior reviewer specifically prepare to lead the discussion of the draft translations in advance. Prior to the meeting, this reviewer makes notes on points of difficulty across translations or in the source questionnaire and reviews translators' comments on their translations and the source documents with a view to managing.

- Ask other team members to review all the draft materials and take notes in preparation for the meeting. The time spent on preparation will be of benefit at the meeting.

- Have the senior reviewer lead the discussion.
  - The lead person establishes the rules of the review process.
  - He/she emphasizes, for example, that most likely the team will change existing translations, and that the common aim is to collaborate towards finding the best solutions.

- Have the senior reviewer appoint two revision meeting note-takers (any careful and clear note-taker with the appropriate language skills, and often the senior reviewer).

- Have the team go through each question, answer scale, instruction, and any other components, comparing draft suggestions, and considering other alternatives. Members aim to identify weaknesses and strengths of proposed translations and any issues that arise such as comparability with the source text, adaptations needed, difficulties in the source text, etc.

- Ensure that changes made in one section are also made, where necessary, in other places. Some part of this may be more easily made after the review meeting on the basis of notes taken.

- Whenever possible, finalize a version for adjudication.
  - If a version for adjudication cannot be produced, the review meeting documentation should note problems preventing resolution.

- After review, before adjudication, copyedit the reviewed version in terms of its own accuracy (consistency, spelling, grammar, etc.).

- After review, before adjudication, copyedit the reviewed version against the source questionnaire, checking for any omissions, incorrect filtering or instructions, reversed order items in a battery or answer scale labels, etc.

Lessons learned

- Guidelines are only as good are their implementation. Quality monitoring plays an essential role. However, evaluation of survey quality begs many issues. Translators asked to assess other translators' work may, for example, be hesitant to criticize or, if not, may apply standards which work in other fields but are not appropriate for survey translation. In the worst instance, they may follow criteria required by people who do not understand survey translation.

- Much remains to be established with regard to survey translation quality. Group dynamics are important. The lead person/senior reviewer leads the discussion. When two suggested versions are equally good, it is helpful to take up one person's suggestion one time and another person's the next time. Given the objectives of the review, however, translation quality obviously takes priority in making decisions about which version to accept.

- Time-keeping is important. The senior reviewer should confirm the duration of the meeting at the start and pace progress throughout. Otherwise much time may be spent on early questions, leaving too little for later parts of the questionnaire.

- It is better to end a meeting when team members are tired and re-convene than to review later parts of the questionnaire with less concentration.

- Practice taking documentation notes on points not yet resolved or on compromised solutions. (See Finding, Selecting, and Briefing Translation Team Members).

- Not everyone needs to be present for all of a review meeting. Members should be called upon as needed. Queries for substantive experts, for example, might be collected across the instrument and discussed with the relevant expert(s) in one concentrated sitting.

6. Have the adjudicator sign-off on the final version for pretesting.

Rationale

Official approval may simply be part of the required procedure, but it also emphasizes the importance of this step and the significance of translation procedures in the project.

Procedural steps

- If the adjudicator has all the skills needed (strong language ability in the source language and target language, knowledge of the study and also survey measurement and design issues), have him or her take part in the review session if this is possible. Even in this case, whenever possible it is advisable to delay official signing-off to another day, thus leaving time for final checking of the decisions taken [12].

- If the adjudicator does not have special relevant expertise, have him or her work with consultants to check that all the procedures have been followed, that appropriate people were involved, that documentation was kept, etc., according to procedural requirements. To assess the quality of review outputs, for example, the adjudicator can ask to have a list of all the perceived challenges and request to have concrete examples of these explained.

- If the expertise of the adjudicator lies somewhere between these extremes, consider having him or her review the translation with the senior reviewer on the basis of the review meeting documentation.

- Ensure again that changes made in one section are also made, if necessary, in other places.

Lessons learned

- Emphasizing the value of finding mistakes at any stage in the production is useful. At the same time, a team effort usually shares responsibility. If things are missed, it is best in any instance if no one is made to feel solely responsible.

- If a translation mistake means a question is excluded from analysis in a national study, the costs and consequences are high; in a comparative survey, the costs and consequences are even higher. Making team members aware of this may help focus attention. The German mistranslation in a 1985 ISSP question regarding participation in demonstrations meant both the German and the Austrian data on this question could not be compared with other countries [3]. (Austria had used the German translation, complete with the mistranslation.)

7. **Pretest the version resulting from adjudication**.

Rationale

One purpose of pretesting is to test the viability of the translation and to inform its refinement, as necessary, in preparation for final fielding.
All instruments should be pretested before use. The best possible version achievable by the team development process should be targeted before pretesting **(**see Pretesting**).**

Procedural Steps

- See Pretesting.

Lessons Learned

- No matter how good the team translation, the review, and adjudication are, pretesting is likely to find weaknesses in design and/or translation [20].

8. Review, revise, and re-adjudicate the translation on the basis of pretesting results.

Rationale

Pretesting results may show that changes to the translation are needed. Changes can be implemented as described below.

Procedural steps

- Decide on the team required to develop revisions. This will differ depending on the nature and number of problems emerging from the pretest and on whether or not solutions are presented along with the problems.

- If a one- or two-person team is chosen that does not include one of the translators, share any changes (tracked or highlighted) with a translator and a “typical target population person” for final commentary, explaining the purpose of the revision.

- Review the documentation from the pretest, considering comments for each question or element concerned.

- Ensure that changes made in one section are also made, where necessary, in other places.

- Copyedit the revised version in terms of its own accuracy (consistency, spelling, grammar, etc.). Target language competence is required for this.

- Copyedit the revised version in its final form against the source questionnaire, checking for any omissions, incorrect filtering or instructions, reversed order items or answer scale labels, etc. Competence in both target and source language is required for this.

- Check in programmed applications that hidden instructions have also undergone this double copyediting (see Instrument Technical Design).

- Present the revised version for final adjudication. The adjudication procedures for this are as before. Project specifics will determine in part who is involved in the final adjudication.

Lessons learned

- It is extremely easy to overlook mistakes in translations and in copyediting. The review and adjudication steps offer repeated appraisals which help combat this, as do the documentation tools.

- It is often harder to find certain kinds of mistakes if one is familiar with the text. It is better if the copyeditors are not the people who produced the texts.

- Although copyediting is a learnable skill, good copyeditors must also have a talent for noticing small details. The senior reviewer should ensure people selected for copyediting work have this ability.

- If the people available to copyedit have helped produce the translations, allow time to elapse between their producing the translation and carrying out copyediting. Even a few days may suffice.

- Problems with incorrect instructions, numbering, filters, and omitted questions are quite common. They are often the result of poor copyediting, cut and paste errors, or inadvertent omissions, rather than "wrong" translation. Thus, for example, reversed presentation of answer scale categories is a matter of order rather than a matter of translation. It can be picked up in checking, even if the reversal may have occurred during translation.

- Use a system of checking-off (ticking) material that has itself been tested for efficiency and usability. In iterative procedures such as review and revision, this checking-off of achieved milestones and versions and the assignment of unambiguous names to versions reduces the likelihood of confusing a preliminary review/adjudication with a final one.

- Automatic copyediting with Word will not discover typographical errors such as for/fro, form/from, and if/of/off. Manual checking is necessary.

9. **Organize survey translation work within a quality assurance and quality control framework and document the entire process.**

Rationale

Defining the procedures used and the protocol followed in terms of how these can enhance the translation refinement process and the ultimate translation product is the most certain way to achieve the translation desired. Full documentation is necessary for internal and external quality assessment. At the same time, strong procedures and protocols do not resolve the question of what benchmarks should be applied for quality survey translation. [6] discusses the need for research in this area.

Procedural steps

The steps involved in organizing a team translation are not repeated here. The focus instead is on what can be targeted in terms of translation quality.

- Define survey translation quality in terms of fitness for use:
    - Fitness for use with the target population.
    - Fitness for use in terms of comparability with the source questionnaire.
    - Fitness for use in terms of producing comparable data (avoiding measurement error related to the translation).
    - Fitness in terms of production method and documentation.

- Produce survey translations in a manner that adequately and efficiently documents the translation process and the products for any users of the documentation at any required stage in production (review, version production control, language harmonization, later questionnaire design).

Lessons Learned

- The effort required to implement a well-structured and well-documented procedure and process will be repaid by the transparency and quality control options it makes possible. Thus even simple Word or Excel templates make it easier to track the development of translations, to check that certain elements have not been missed, and to verify if and how certain problems have been resolved. These might begin with translator notes from the draft productions and evolve into aligned translations in templates for review, later becoming templates for adjudication with translations proposed and comments on these. [2] provides examples of how Excel templates help guide quality control and assurance steps.

- Once procedures become familiar and people gain practice in following protocols, the effort involved to produce documentation is reduced.

Appendix A

Documentation templates

Template 1 is typical of templates used in the European Social Survey (ESS) in rounds 1–4 for draft translations. The source questionnaire has been entered in the template in distinct sections. Each translator enters his or her translation in the template and provides commentary. For later stages in the translation process, similar templates retained information from each preceding stage and added columns for outcomes and comments on the current step (see Template 2).

Template 1: Extract from a translation template from the ESS Round 4 for one draft translation (core module B)

	Source English Section B	**Routing**	**Draft Translation 1**	**Comments**
B above B1	**Now we want to ask a few questions about politics and government**			
B1	How interested would you say you are in politics – are you…			
I in B1	**READ OUT…**			
I in B4-B10	**READ OUT…**			
B4	…[country]'s parliament?			
B5	…the legal system?			
B6	…the police?			
B7	…politicians?			
B8	…political parties?			
B9	…the European Parliament?			
B10	…the United Nations?			
RC, B4-B10	**No trust at all**			
	Complete trust			
	(Don't know)			

B = Bridge; CI = Coding / Design Instruction; I = Interviewer Instruction; RC = Response Category; RI = Respondent Instruction

Template 2 illustrates possible headings for a template bringing together two draft translations for a review meeting based on Template 1.

Template 2: Headings required for a team review meeting

	Source English Section B	Routing	Draft Translation 1	Comments	Draft Translation 2	Comments	Review version	Comments from review meeting
B above B1	**Now we want to ask a few questions about politics and government**							
B1	How interested would you say you are in politics – are you…							
I in B1	**READ OUT…**							
RC	very interested,							
	quite interested,							
	hardly interested,							
	or, not at all interested?							
	(Don't know)							
B2	**CARD 6**							
	How often does politics seem so complicated that you can't really understand what is going on?							
	Please use this card.							
RC	Never							
	Seldom							
	Occasionally							
	Regularly							
	Frequently							
	(Don't know)							

B = Bridge; CI = Coding / Design Instruction; I = Interviewer Instruction; RC = Response Category; RI = Respondent Instruction

Appendix B

Tasks of personnel in team translation projects

Translators

- Prepare individual translations in preparation for the review session.
- Take notes on translation and source texts in preparation for the review session (documentation to inform the review).
- Participate in review sessions with other members of the review team.
- Consult on any translation revisions at later stages.
- May assess source questionnaires for comparative viability.
- May assess other survey translations.
- May assist in copyediting.

Reviewers

- Participate in review sessions at times identified as relevant depending on their role.
- Contribute their individual area of expertise to developing and refining the translated instrument.

Senior reviewer

- Organize review session meetings (unless a co-coordinator does this).
- Organize materials for the review session(s) (unless a co-coordinator does this).
- Lead review sessions, including attending to group dynamics, appointing note takers, coordinating contributions to the discussion, ensuring the meeting runs according to schedule, and ensuring each relevant topic is discussed and resolved or noted as unresolved.
- Organize and supervise the documentation of review session outputs. Review session outputs will principally consist of refined translation versions and accompanying documentation, queries, and comments; they may also include action points arising from the review meeting(s), such as the need to consult with question designers or other subject matter experts.

Adjudicator

- Appraise and officially sign off on translations, usually after the review meeting(s).
- Appraise the review outputs probably in consultation with a senior advisor (the senior reviewer or other consultant) and approve a final version for pretesting and fielding. If the adjudicator is also the senior reviewer, review and adjudication may follow directly upon one another.
- If the senior person on a project who is officially required to sign off on a translation is not appropriate to appraise translation

quality and decisions, this nominal adjudicator may delegate adjudication to another senior person better suited for this task. Alternatively, in the same situation, the adjudicator may use consultants and documentation from the review session(s), to work through the translation and documented decision points and notes before signing off.

Copyeditor(s)

- Check for correctness in the target language, including spelling, omissions, wrong formatting, consistency of formulation, and repeated phrases (e.g., "please tick one box"), and for completeness of revision. When multiple versions are in circulation, teams can become unclear, for example, about which version is indeed intended to be the final version. Copyeditors should also check this.
- Check against the source document for such errors as inadvertent omissions or additions or question and answer option reversals, mistakes resulting from copy-and-paste activities, misread source questions, and filter numbering correctness.

Co-ordinator

- Large translation efforts, centrally organized studies, or efforts conducted within a large organization may have a coordinator to manage the translation effort in an organizational management sense (schedule coordination, personnel identification, budgeting, and so forth).
- In other instances the senior reviewer may organize the translation effort.

Substantive and other experts

- Substantive experts may be needed to provide advice on a variety of matters, such as the suitability of indicators or the formulation of questions with regard to measurement goals.
- Question design experts might be consulted about changes in format necessitated by translation.
- Interviewers might be consulted for fielding matters relevant to translation.
- Visual design experts might, for example, be consulted about cross-cultural aspects of visual presentation.

Programmers

- If the questionnaire is computer-assisted, consultation with programmers, or those familiar with programming requirements, is needed to ensure that the translation document or file is marked appropriately. Numerous programming details may need to differ from one language to another to accommodate different language structure requirements (see Questionnaire Design).

Back-up personnel

- Projects sometimes run beyond agreed times of availability of personnel. Personnel may also become unavailable for a variety of reasons. It is a good idea to have back-up personnel in place.

External assessors

- If some parts of the translation process or translation outputs are to be subjected to external assessment, suitable assessment personnel will be required (see Translation Assessment).

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).
Adjudicator	The person who signs-off on a finalized version of a questionnaire (see Adjudication).
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).

Consistency	Consistency is achieved when the same term or phrase is used throughout a translation to refer to an object or an entity referred to with one term or phrase in the source text. In many cases, consistency is most important with regard to technical terminology or to standard repeated components of a questionnaire. Reference to "showcard" in a source questionnaire should be consistently translated, for example. The translation of instructions which are repeated in the source text should also be repeated (and not varied) in the target text.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Copyeditor	The person who reviews a text and marks up any changes required to correct style, punctuation, spelling, and grammar errors. In many instances, the copyeditor may also make the corrections needed.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Full translation (double translation)	Each translator translates all of the material to be translated. It stands in contrast to split translations.
Indicator	The third step in the concept/construct/indicator/question model. They relate to behaviors, attitudes, reported facts, etc., considered to provide indirect measurement of constructs. Several indicators might be used for a given construct. For example, price, durability, the attractiveness of packaging, and purchasing convenience (ease), can be indicators to measure a construct centered on customer satisfaction with a given product.

Item	Researchers differ greatly in how they use this term. It is usually and most correctly used to refer to the statements in Likert-type batteries. Example: *The Government should provide jobs for everyone who wants to work.*
Language harmonization	Language harmonization can be understood as the procedures and result of trying to find a common version (vocabulary and/or structure) across questionnaires for different regional varieties of a “shared” language.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Mode	Method of data collection.
Overlap in the split translations	A compromise solution between split and full translations is to ensure that some overlap exists between materials divided among translators. The material is split up the way cards are dealt in many games, everyone getting a spread of the material. Each translator could then receive the last one or two questions of another translators "piece." This allows the review team members to have an increased sense of whether differences in translating approaches between translators and their understanding of source text components at the draft production level.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.

Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Reviewer	Person who participates in the review of translations in order to produce a final version (see Appendix A of Translation).
Source document	The original document from which other (target) documents are translated or adapted as necessary.
Source language	The language in which a questionnaire is available from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Source questionnaire	The questionnaire taken as the text for translation.
Split translation	Each translator translates only a part of the total material to be translated in preparation for a review meeting, in contrast to translating the entire text (see full translation).
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.

Team translation Team approaches to survey translation and translation assessment bring together a group of people with different talents and functions in the team so as to ensure the mix of skills and discipline expertise needed to produce an optimal translation version in the survey context. Each stage of the team translation process builds on the foregoing steps and uses the documentation required for the previous step to inform the next. In addition, each phase of translation engages the appropriate personnel for that particular activity and provides them with relevant tools for the work at hand.

Total Survey Error (TSE) Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.

Translator The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other tasks such as reviewing and copyediting.

Variance A measure of how much a statistic varies around its mean over all conceptual trials.

References

[1] Acquadro, C., Jambon, B., Ellis, D., & Marquis, P. (1996). Language and translation issues. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trial* (2nd ed.) (pp. 575–585). Philadelphia, PA: Lippincott-Raven.

[2] Dept, S., Ferrari, A., & Wäyrynen, L. (2008, June). *Comparative overview of test adaptation and translation verification procedures in three international surveys*. Paper presented at the international conference on survey methods in multinational, multiregional, and multicultural contexts (3MC), Berlin, Germany. Retrieved April 16, from http://www.csdiworkshop.org/pdf/3mc2008_proceedings/session_09/Dept.pdf

[3] Harkness, J. A. (2010, March). *Survey translation and the question of quality*. Paper presented at the International Workshop on Comparative Survey Design and Implementation, Lausanne, Switzerland.

[4] Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. D. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 56–77). New York/London: Lawrence Erlbaum Associates.

[5] Harkness, J. A. (2000/2008). *Round 4 ESS translation strategies and procedures*. Retrieved April 16, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_docman&task=doc_download&gid=351&itemid=80

[6] Harkness, J. A. (2007). Improving the comparability of translation. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European Social Survey* (pp. 79–94). London: Sage Publications.

[7] Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. van de Vijver & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: John Wiley & Sons.

[8] Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R., & Villar, A. (2010). Designing questionnaires for multipopulation research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E., Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural, and multiregional contexts* (pp. 33–58). Hoboken, NJ: John Wiley & Sons.

[9] Harkness, J. A., Pennell, B-E., & Schoua-Glusberg, A. (2004). Questionnaire translation and assessment. In S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin & J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*

(pp. 453–473). Hoboken, NJ: John Wiley & Sons.

[10] Harkness, J. A., Schoebi, N., Joye, D., Mohler, P., Faass, T., & Behr, D. (2008). Oral translation in telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas et al. (Eds.), *Advances in telephone survey methodology* (pp. 231–249). Hoboken, NJ: John Wiley & Sons.

[11] Harkness, J. A., & Schoua-Glusberg, A. S. (1998). Questionnaires in translation. In J. A. Harkness (Ed.), *Cross-cultural survey equivalence* (ZUMA-Nachrichten Spezial 3) (pp. 87-126). Mannheim, Germany: ZUMA.

[12] Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 117-140). Hoboken, NJ: John Wiley & Sons.

[13] Harkness, J., Villar, A., Kephart, K., Behr, D., & Schoua-Glusberg, A. (2009, March). *Research on translation assessment procedures: Back translation and expert review.* Paper presented at the International Workshop on Comparative Survey Design and Implementation, Ann Arbor, MI.

[14] Harkness, J. A., Villar, A., Kephart, K., Schoua-Glusberg, A, & Behr. D. (2009, May). *Survey translation evaluation: Back translation versus expert review.* Paper presented at the annual conference of the American Association for Public Opinion Research, Hollywood, FL.

[15] Harkness, J. A., Villar, A., Kruse, Y., Branden, L., Edwards, B., Steele, C., & Wang, Y. (2009, March). *Interpreted telephone survey interviews.* Paper presented at the International Workshop on Comparative Survey Design and Implementation, Ann Arbor, MI.

[16] Harkness, J. A, Villar, A., Kruse, Y., Steele, C., Wang, Y., Branden, L., Edwards, B., & Wilson, C. (2009, May). *Using interpreters in telephone surveys.* Paper presented at the annual conference of the American Association for Public Opinion Research, Hollywood, FL.

[17] McKay, R. B., Breslow, M. J., Sangster, R. J., Gabbard, S. M., Reynolds, R. W., Nakamoto, J. M. et al. (1996). Translating survey questionnaires: Lessons learned. *New Directions for Evaluation, 70* (Summer), 93–105.

[18] Pan, Y., & de la Puente, M. (2005). Census Bureau guidelines for the translation of data collection instruments and supporting materials: Documentation on how the guideline was developed. *Statistical Research Division's Research Report Series* (Survey Methodology

#2005-06). Washington, DC: U.S. Bureau of the Census.

[19] Schoua-Glusberg, A. (1992). *Report on the translation of the questionnaire for the national treatment improvement evaluation study*. Chicago, IL: National Opinion Research Center.

[20] Willis, G. B., Kudela, M. S., Levin, K. L., Norberg, A., Stark, D. S., & Forsyth, B. H. et al. (2010). Evaluation of a multistep survey translation process. In J. A. Harkness., M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 137–152). Hoboken, NJ: John Wiley & Sons.

Internet Links

[21] Afrobarometer Survey. Retrieved July 10, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[22] Asian Barometer. Retrieved July 10, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

[23] European Social Survey (ESS). Retrieved July 10, 2010, from http://www.europeansocialsurvey.org

[24] Survey of Health, Ageing and Retirement in Europe (SHARE). (2010). Retrieved July 10, 2010, from http://www.share-project.org/

VIII. Translation: Translation Assessment

Janet Harkness

Under Development!

The chapter on translation assessment will consider different forms of qualitative and quantitative assessment related to translation and present the current state of research and relevant literature as available. The material will be divided into subsections as follows:

- Assessment and survey translation quality

  Assessment and evaluation assume that criteria of evaluation are available with which to assess the quality of given translation products and benchmarks and that standards exist against which translation products can be "measured". In the survey research field there is only limited consensus on what these criteria and benchmarks might be and what translations that meet these criteria might then look like. This section will deal with these issues. It will identify criteria of obvious relevance for survey translations and will identify others which may or may not be of relevance in a given context.

- Assessment as part of team translation

  Qualitative assessment of translation drafts as they are being developed is an integral and essential component of team translation procedures as described in the translation guidelines (see Translation). This section will identify these steps and the (partially iterative) procedures involved in each.

- Assessment using external translation assessors

  This section will describe various models of how external reviewers can be used in survey translation efforts. Some projects currently rely on external review teams to provide most of their assessment; others combine internal assessment procedures with outside quality monitoring.

- Assessment using focus groups and cognitive interviews

  This section will outline how various pretesting methods can be used to gain insight the appropriateness of language used in survey translations and will indicate recent developments in procedures.

- Assessment using quantitative analyses

  Quantitative assessment procedures are also used to assess whether translated instruments perform as expected or hoped. This section will describe what qualitative assessment can add to qualitative assessments and will outline some of the main procedures used. These include:

- Multi Trait Multi Method
- Item response theory (IRT) and differential item functioning (dif)
- Various form of split ballot tests.

VIII. Translation: Language Harmonization

Janet Harkness

Introduction

Language harmonization is developing a common version (vocabulary or structure) across questionnaires for different regional varieties of a “shared” language. The guidelines in this chapter address the fact that it is important for countries or locations that share a language to take steps to avoid unnecessary differences across their questionnaires [5] [6] [8].

Why harmonize language?

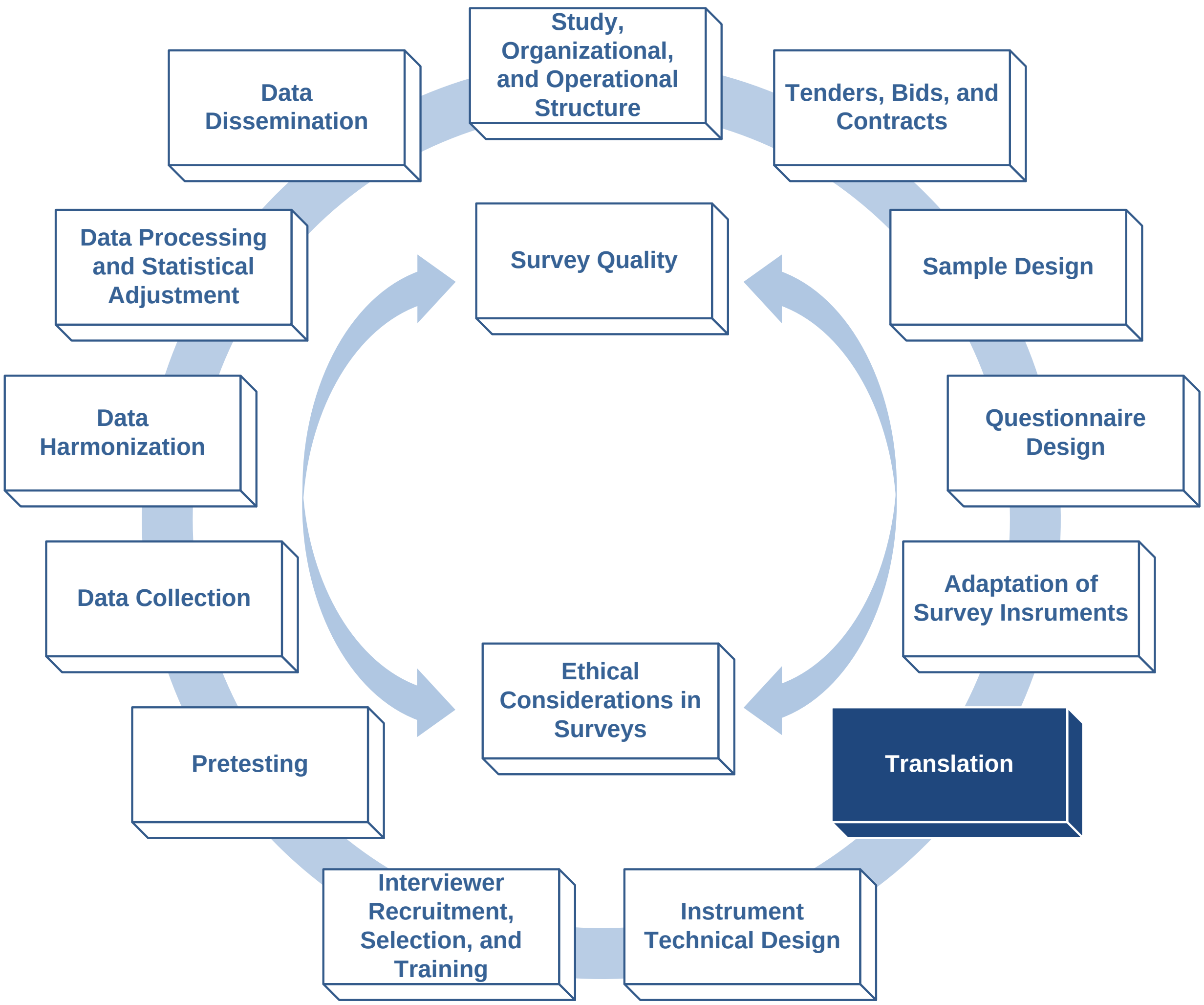
In cross-national surveys, multiple countries or communities may field surveys in the same language. Languages such as Russian, French, German, Spanish, and Chinese, for example, are spoken as a first language by populations in a number of countries. However, the regional standard variety of a language used in one country usually differs to varying degrees in vocabulary and structure from regional standard varieties of the same language used in other countries. For example, American English, British English, and Indian English differ systematically in a variety of ways. Often differences relate to vocabulary and pronunciation, but differences in syntax and other grammatical features of the language are also found.

As a result, translations produced in different locations may differ considerably—not only because there is usually more than one way to translate a question (see Translation) but because of regional differences in language, social reality, and culture. Thus differences in translation may reflect the given regional standard (e.g., Mexican Spanish versus Castilian Spanish), may simply reflect the fact that there is more than one way to say and to translate the same source text, may actually reflect different interpretations of what the source text intends to convey, or may stem from different social and cultural realities.

A further complicating factor is that the written regional standard variety of a language may differ systematically and markedly from the spoken form of that language the same community uses. Spoken Swiss German, for example, differs notably from region to region. However, each spoken variety also differs markedly and in some shared ways from standard written Swiss German. Standard written Swiss German, in turn, differs in some limited respects from the standard written forms of German used in Germany and Austria [3] [9]. Research on how interview scripts for such contexts might best be developed is in its infancy.

As described below, language harmonization usefully takes place at a late stage of translation. Figure 1 shows translation within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

1. **Harmonize the wording of questionnaires in one language whenever possible.**

   Rationale

   All else being equal, it is preferable to keep the wording constant within a language across locations. If no policy of harmonization is followed, unnecessary differences may proliferate. Some of these, such as differences in translating answer scales, may negatively affect measurement [13].

Procedural steps

- Decide upon the policy and procedures to be adopted on harmonization (obligatory or optional, full or optimized, top-down or bottom-up; see Appendix A).
- Decide on the tools to be used; these should include a documentation component.
- Inform all locations sharing a language of the harmonization policy and procedures and related requirements.
- Schedule and organize any translations so that harmonization is possible.
- If working from a single translated questionnaire towards localized versions, prepare and distribute the single translation. If such a top-down approach is used, the single translated version should be produced in a team translation approach that includes input for the different regional varieties of the languages that are to be accommodated.

Lessons learned

- The increased effort, time, and outlay to undertake harmonization may be an obstacle to implementing it.
- Without advance planning, the short time often available for translation may make harmonization preparation and meetings to discuss versions difficult and makes pretesting of alternatives unlikely.
- Without clearly defined protocols and some training, the local teams asked to harmonize may have difficulty making informed decisions about harmonization. They may also not properly record their decisions and their motivations.
- When new locations join an ongoing study, new harmonization needs may arise in previously harmonized versions of questions. No research could be identified on whether it is better for the older harmonization decisions to be kept and the new country to deviate or for all to change. There is "received wisdom" about changing as little as possible but this is always over-ruled when change becomes necessary.
- Content management system and localization software can aid

identification of text requiring harmonization and provide a documentation option for differences retained (see Translation Tools).

- Keeping the words the same across questionnaires in different locations does not automatically mean that perceived meaning and intended measurement are retained across populations. Pragmatic meaning also needs to be considered (see [2] [7] [10] [12]). At the same time, there is little research available that clarifies how to keep both semantic meaning and pragmatic meaning stable across surveys in different languages. Pragmatic considerations might also stand in conflict with retaining semantic meaning. It remains to be established how "sameness" and comparability are best ascertained at the textual level (see [2] [4] [7]).

- Localized versions based on a single common translation may have fewer differences across versions in a shared language. This does not mean that the instruments are necessarily better than those with more differences. Careful testing should be carried out to make sure that each population does understand the questions as intended [10] [11].

2. Only keep necessary differences.

Rationale

There are often several ways to formulate a survey question, an explanation, or even instructions. Teams cooperating in a harmonizing effort must try to lay aside personal preferences. Differences that are maintained across questionnaires should be considered genuinely necessary—and, preferably, demonstrated through testing to be so.

Procedural steps

If harmonization takes place on the basis of individual draft translations made by each national or regional group (bottom-up approach):

- Organize templates to enable easy comparison of the draft translations to be reviewed for harmonization.

- Organize the harmonization meeting(s).
  - These can be face-to-face, perhaps piggy-backing on another meeting. However, webcasting, webinars, or “skyping” may be the only affordable modes of meeting.
  - Share versions prior to the meeting and produce a central document aligning them side by side; use a format that also allows each user to see the source and target questions easily (see Appendix B).

    - If possible, appoint someone to identify types of difference (or just differences) ahead of the meeting, both on the basis of any past experience and by checking the draft versions to be harmonized. If this person is someone who also attends the meeting, he or she might usefully introduce each question, summarizing points noticed.
    - Appoint a meeting chair and determine how group/location decisions will be made, ensuring fair representation of each group/ location.

If common wording in the form of a single translated version is the starting point (top-down approach):

- Organize templates to enable easy comparison of the suggested localizations.

- Have each local team propose modifications it considers necessary to the common version.

- Organize the reconciliation meeting(s).
    - These can be face-to-face if possible, perhaps piggy-backing on another meeting. However, webcasting, webinars, or "skyping" may be the only affordable modes of meeting.
    - Define the goals of this meeting (e.g., to review suggested changes, to try to find new shared alternatives, to share questions about the single translation).
    - Share localization suggestions prior to the meeting and produce a central document aligning them side by side; use a format that also allows the users to see the source questions easily.
    - If possible, appoint someone to identify the types of localization proposed ahead of the meeting, both on the basis of any past experience and by checking the localizations proposed. If this person is someone who also attends the meeting, he or she might usefully introduce each question, summarizing the suggestions made and questions raised.
    - Appoint a meeting chair and determine how decisions will be made, ensuring a fair representation of each group/location.

Lessons learned

- Personal language perception and usage can be mistaken for generic language usage. It would be mistaken to assume that because one or more speakers make a distinction that these are then distinctions made by all speakers of a given speech community.

- It may not serve the study's purpose to make decisions on the principle of a "majority" vote. The aim is ultimately to allow necessary difference in any given version.

- Harmonization is not limited to the choice of words or phrases; it can include decisions, for example, about how sentences are structured and answer scales organized.

- Sometimes harmonization takes the form of adding a term or an example to whatever is common with other shared language versions. Thus if a question about tobacco use does not cover a special form that is only relevant (but important) for one population, mention of it could be added for that population alongside the other forms of tobacco use mentioned in the other versions of the question. This strategy of keeping what is common but adding a local requirement is frequently found in adaptations (see Adaptation of Survey Instruments).

- If the top-down localization model is used, teams may spend more time discussing the single translation than any of their localizations. This has advantages and disadvantages. One benefit in discussing the available translation is that the group may have new ideas about a possible common version or a common version with occasional "add-ons" as just described. One possible disadvantage is that consideration of the range of localized suggestions is reduced, with each team members ultimately focusing more on resolving what to choose for his or her own version.

3. Schedule harmonization at an appropriate time.

Rationale

Harmonization efforts can result in changes in one or all questionnaires. The harmonization decisions need to be made when each questionnaire version (or the single translation) is at an advanced stage of development. Although desirable, iterative rounds of pretesting are not likely to be feasible. Thus if a team translation procedure (documented translation review, adjudication, and pretesting) is followed, harmonization should precede pretesting and thus final adjudication (see Translation and Appendix A). Pretesting can be used to check harmonization decisions. It may also indicate that further changes are required in one or more versions.

Procedural steps

- Identify the time at which a well-developed version of each questionnaire to be harmonized will be pretested (or the single common version is well advanced) and arrange for harmonization before that time.

Lessons learned

- If countries are fielding at different times, a group fielding much later than others may have trouble carrying out (or funding) harmonization preparations in time for groups fielding earlier. The sooner harmonization is organized and scheduled, the greater the chances are of successful schedule coordination between countries or locations.

- In practice, recommending harmonization rather than requiring it may not be sufficient to motivate countries or locations to engage in the extra effort. The European Social Survey (ESS) aimed for optimized harmonization and recommended it to participating countries. Since harmonization was not a requirement, countries were left with considerable freedom as to whether they harmonized or not. The countries' various time schedules also did not easily accommodate a harmonization step. Harmonization was further complicated by countries with shared languages joining the project at different times. For example, the first Russian translation was produced in Israel for Russian-speaking immigrants there; the questionnaire for Russians in Russia was only produced several years later [1].

- Without harmonization, the differences that may arise across different regional versions of questionnaires in a shared language can be considerable and may often be unnecessary [1] [5].

- The differences in regional varieties of languages, at least in terms of what needs to be captured in questionnaires, may sometimes also be overestimated.

- While recognizing and emphasizing that same wording does not mean same meaning or comparable measurement, differences across questionnaires may introduce unnecessary and potentially serious measurement error. It is, therefore, important to include harmonization procedures in the study design.

4. **Determine and stipulate documentation requirements and tools for the process and outcomes.**

Rationale

Those undertaking documentation should have a clear understanding of what is required and should be provided with aids that enable them to maintain documentation without undue burden. Documentation templates play an essential role while deliberating on harmonization as described above. Documentation also provides the evidence examined in quality monitoring and assurance steps, for any coordination of harmonization efforts that may exist in a project and provides secondary analysts and other users of data with information about differences across instruments.

Procedural steps

- Determine documentation needs and create stipulations to be followed by those involved in harmonization in order to achieve these needs.
- Develop templates for the language harmonization process and the harmonization outcomes (see Appendix B).
- Distribute templates and specifications to all involved well in advance and ensure they are familiar with their purpose and how to use them.
- Provide examples of what is sufficient documentation and what is not.

Lessons learned

- Good and accessible documentation is essential to language harmonization efforts. It enables teams to compare options more easily while making decisions and also to record clearly the decisions taken. Users of data also benefit from documentation on differences across instruments.

5. Undertake language harmonization within a quality assurance and control framework as that relates to translation quality.

Rationale

Language harmonization is undertaken to reduce unnecessary variance across versions of a questionnaire in one language that may negatively affect measurement in any of a variety of ways. The purpose of harmonization is, thus, to enhance measurement quality.

Procedural steps

- Plan and undertake harmonization in controlled procedures as described above.
- Plan to follow harmonization with a pretesting phase.
- Develop the relevant materials needed as described above.
- Identify and engage suitable people to be involved in harmonization as described above.
- Brief team members on the materials, purpose and strategies used in harmonization.

- Complete the main harmonization process.
- Pretest and then modify instruments as relevant.
- Share findings in a well-documented and timely fashion with any coordinating center, as relevant.

Lessons learned

- The more rigorous the translation procedures and the various sub-activities such as harmonization and pretesting become, the more important scheduling, budgeting, and briefing are.
- Long-term, the benefits of having and being able to share well-developed and tested instruments can be very considerable.
- It is more effective to require locations to engage in harmonization than to recommend that they do.

Appendix A

Ways to organize and implement language harmonization

There are several ways to organize and implement harmonization with regard to whether it is obligatory or not and in terms of how the procedure is organized. These are outlined in Table 1 below.

Table 1: Language harmonization options

Term	Explanation	Advantages	Disadvantages
Obligatory harmonization	The project stipulates that language harmonization (in whatever form) must be undertaken.	Participating locations will be more likely to engage in harmonization procedures.	Obligatory participation might be a real burden on some participants.
		Unnecessary differences have a better chance of being avoided.	Group dynamics can sometimes mean that one location dominates discussion. For example, a location using the language as a first language for the majority of its population may seek or be given more influence in the discussion than a location using the language for a minority of its population.
Optional harmonization	The project recommends harmonization but does not make it an obligatory requirement.	Recommending rather than requiring harmonization might be a more realistic requirement in some contexts.	A recommendation may not be enough to ensure countries engage in the additional effort required.
			Unnecessary differences across versions and negative effects on measurement may result.
Full harmonization	The project aims to produce a single language version to	The wording of the questions is the same in each location.	The "same" wording may be systematically understood differently

Term	Explanation	Advantages	Disadvantages
	be used for all the locations using that language.		in different locations, or not understood in one or more locations.
Optimized harmonization	The project aims to harmonize as much as possible but to permit local divergence from the shared wording as necessary. Harmonization is pursued only to the degree to which it optimizes comparability.	As much as possible is kept common but needed differences are permitted.	Teams may have difficulty distinguishing between their preferences and what are really required differences. This holds for bottom-up and top-down approaches.
			Teams will usually lack experience in harmonization decision-making. This holds for bottom-up and top-down approaches.
Top-down approach (localization from single version)	A single target language version is first produced. This is then adjusted as necessary for the different varieties of the target language. Production of the single version should take into consideration the needs of the different language varieties to be accommodated. The team translation procedures described in Translation would be useful for this.	By beginning with a shared common version, locations may end up with more shared common wording than by using a bottom-up approach.	The success of the single translation in anticipating and accommodating needs of different locations can determine how much of the translation is left intact. If the single translation meets with opposition from many groups/locations involved with respect to many components, this will greatly complicate the harmonization effort.
		Teams will usually lack experience in harmonization decision-making. This holds for bottom-up and top-down approaches	The fact that one translation (and only one) is on the table may make it harder to spot where differences are needed.
		.	People might not propose alternatives they would have seen if each location had made an independent translation.
			Shared wording might

Term	Explanation	Advantages	Disadvantages
			not mean shared understanding or comparable measurement
Bottom-up approach (harmonization of different versions)	Each location produces a draft translation. A good version produced on the basis of team translation prior to pretesting should suffice (see Translation). These translations form the basis of the harmonization review.	Every location has studied the source questionnaire and considered an optimal version for their location.	Locations may be unwilling to produce a draft translation that is ultimately changed again.
		The harmonization review has all the alternatives at its disposal to decide commonalities, possibly find new shared language and determine and document needed differences.	Locations might over-perceive the need to retain their versions.
			Teams may have difficulty distinguishing between their preferences and what are really required differences. This holds for bottom-up and top-down approaches.
			Teams will usually lack experience in harmonization decision-making. This holds for bottom-up and top-down approaches.

Appendix B

Documentation templates

Clear instructions and documentation templates help researchers conduct and document harmonization products. Below are a few examples of templates used in recent cross-national surveys in connection with language harmonization.

The WHO Mental Health Survey Initiative (MHSI) aimed for an optimized and maximally harmonized questionnaire. The output of harmonization procedures for Spanish in Latin America and Spain is presented in Table 1 below.

Table 1: Examples of harmonization carried out in Spanish-speaking countries in MHSI.

A	B	C	D
English Term **Término en inglés**	***Terms proposed for Spanish*** **Términos propuestos en español**	***Terms actually chosen*** **Términos seleccionados**	***Terms used in individual locations when harmonization not possible*** **Términos alternativos según país****
Free base, (cocaine-based drug)	Free base		Basuco(1, 3, 8), pasta base (6)
Herbalists	Herbolarios, Naturistas		Naturistas (1,2), homeópatas (1,2), herbolarios (1), herbalistas (2) yerberos/ yerbateros (3, 8)
Hot flashes	Sofocos		Sofocos(1), sofocones (2), bochornos (5,6), calores (8)
Ulcer in your stomach or intestine	Úlcera estomacal o intestinal	Úlcera de estómago o intestinal	
Unhappy	Desdichado(a) Desgraciado(a)	Infeliz o desgraciado(a)	
Upset	Molesto	Alterado	
Using a 0 – 10 scale	Utilizando una escala de 0 a 10	En una escala de 0 – 10	
Usual, usually	Habitual, Habitualmente		Habitual/habitualmente (1), usual/usualmente (2)

A	B	C	D
Normally	Normalmente	Generalmente	
Was it before you were a teenager?	¿Fue antes de la adolescencia?	¿Fue antes de los trece años?	
What is the day of the week?	¿A qué día de la semana estamos?	¿En qué día de la semana estamos?	
What is the longest period of days, weeks, months, or years you were...?	¿Cuánto duró el periodo más largo de días, semanas...?	¿Cuántos días, semanas, meses o años duró el periodo más largo durante el que…?	
What number describes...?	¿Qué cifra describe...?	¿Qué número describe mejor...?	
What season of the year is it?	¿En qué estación…?		¿En qué estación (1), época (3,8), del año estamos?

Note: The numbers in Column D indicate the countries using the term, as follows: (1) Spain, (2) Latin America, (3) Colombia, (4) Puerto Rico, (5) Mexico, (6) Chile, (7) Argentina, (8) Panama. Table 1 is adapted from [8].

The coordinating team on translation in the European Social Survey (ESS) investigated differences across shared language versions in the survey using templates similar to Template 1 below. This template brings together German translations made for different countries and comments on any documentation made in various countries on differences. It was not intended for public use. The people using it understood German and therefore did not explain everything noted to each other. A document for public use would be more explicit.

Template 1: German translations across participating countries

Code	Source	German Austria	German Germany	German Lux	German Switzerland	Comment
Q. A1	**On an average weekday, how much time, in total, do you spend watching television?**	Wie viel Zeit verbringen Sie an einem normalen Wochentag insgesamt mit Fernsehen?	Identical to Lux. Wie viel Zeit verbringen Sie an einem gewöhnlichen Werktag insgesamt damit, fernzusehen?	Identical to Germany	Karte 1. Wie viel Zeit verbringen Sie an einem gewöhnlichen Werktag insgesamt mit Fernsehen?	weekday versus work day: not mentioned in notes Watching TV explicit in D/L (verb formulation) nominalized in A and CH; not commented on
I	**Please use this card to answer.**	Bitte verwenden Sie diese Karte zur Beantwortung.	Bitte sagen Sie es mir anhand von Liste 1.		Bitte verwenden Sie für Ihre Antwort Karte 1.	House styles not commented on
RC	**No time at all**	See GER/Lux gar keine Zeit	See Austria/Lux Gar keine Zeit		Überhaupt keine Zeit	no comments on differences between CH and the others
	Less than ½ hour	See CH weniger als ½ Stunde	Weniger als eine 1/2 Stunde		See Austria Weniger als ½ Stunde	Differences not commented upon
	½ hour to 1 hour	mehr als ½ Stunde, bis zu 1 Stunde	1/2 bis zu 1 Stunde		½ Stunde, bis zu 1 Stunde	"More than ½ an hour up to 1 hour "versus "½ to 1 hour" or "½ an hour to 1 hour" CH comma possibly disruptive for reading.

Note: The header "Code" in the first column on the left refers to the abbreviations in that column; QA1 = the question code, I = Instructions, RC = response categories.

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Content management	The software and procedures used to capture, save, organize, and distribute information in digitalized form.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.

Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.

Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.

Team translation	Team approaches to survey translation and translation assessment bring together a group of people with different talents and functions in the team so as to ensure the mix of skills and discipline expertise needed to produce an optimal translation version in the survey context. Each stage of the team translation process builds on the foregoing steps and uses the documentation required for the previous step to inform the next. In addition, each phase of translation engages the appropriate personnel for that particular activity and provides them with relevant tools for the work at hand.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.

References

[1] Andreenkova, A.V. (2008, June). *Translating the instrument for a comparative survey in Russian*. Paper presented at the International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC), Berlin, Germany.

[2] Braun, M., & Harkness, J. A. (2005). Text and context: Challenges to comparability in survey questions. In J. H. P. Hoffmeyer-Zlotnik & J. A. Harkness (Eds.), *Methodological aspects in cross-national research* (ZUMA-Nachrichten Spezial 11) (pp. 95–107). Mannheim, Germany: ZUMA. Retrieved April 29, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial11.pdf

[3] Fishman, J. A. (2003). Bilingualism with and without diglossia; diglossia with and without bilingualism. In C. B. Paulston & G. R.Tucker (Eds.), *Sociolinguistics: The essential readings* (pp.359-366). Oxford: Blackwell.

[4] Harkness, J. A. (2010, March). *Survey translation and the question of quality*. Paper presented at the International Workshop on Comparative Survey Design and Implementation, Lausanne, Switzerland.

[5] Harkness, J. A. (2000/2008). *Round 4 ESS translation strategies and procedures*. Retrieved April 16, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_docman&task=doc_download&gid=351&itemid=80

[6] Harkness, J. A. (2007). Improving the comparability of translation. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European Social Survey* (pp. 79–94). London: Sage Publications.

[7] Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp.19-34). Hoboken, NJ: John Wiley & Sons.

[8] Harkness, J. A., Pennell, B-E., Villar, A., Gebler, N., Aguilar-Gaxiola, S., & Bilgen, I. (2008). Translation procedures and translation assessment in the World Mental Health Survey initiative. In R. Kessler. & T. B. Ustun (Eds.), *The World Health Organization Mental Health Survey* (pp. 91–113). New York, NY: Cambridge University Press.

[9] Harkness, J. A., Schoebi, N., Joye, D., Mohler, P., Faass, T., & Behr, D. (2008). Oral translation in telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas et al. (Eds.), *Advances in telephone survey methodology* (pp. 231–

249). Hoboken, NJ: John Wiley & Sons.

[10] Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 117-140). Hoboken, NJ: John Wiley & Sons.

[11] McKay, R. B., Breslow, M. J., Sangster, R. J., Gabbard, S. M., Reynolds, R. W., Nakamoto, J. M. et al. (1996). Translating survey questionnaires: Lessons learned. *New Directions for Evaluation, 70*(Summer), 93–105.

[12] Schwarz, N., Oyserman, D., & Peytcheva, E. (2010). Cognition, communication, and culture: Implications for the survey response process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp.177-190). Hoboken, NJ: John Wiley & Sons.

[13] Villar, A. (2009). *Agreement answer scale design for multilingual surveys: Effects of translation-related changes in verbal labels on response styles and response distributions*. Unpublished doctoral dissertation, University of Nebraska, Lincoln, NE.

VIII. Translation: Finding, Selecting, and Briefing Translation Team Members

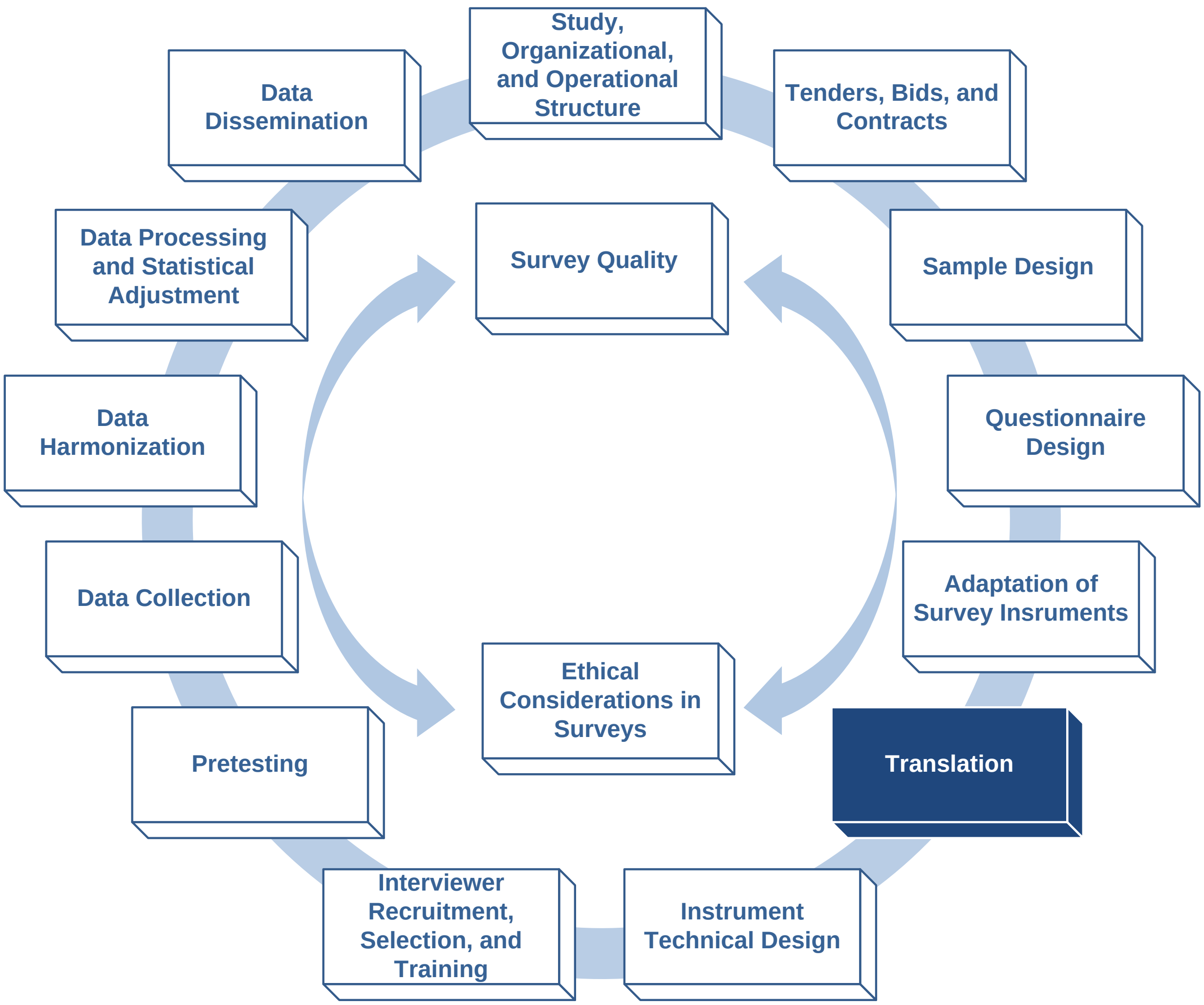
Janet Harkness with (alphabetically) Dorothée Behr, Ipek Bilgen, AnaLucía Córdova Cazar, Lei Huang, An Lui, Mathew Stange, and Ana Villar

Introduction

The following guidelines describe how to find and select suitable people for a team translation effort; they also outline briefing for members of the team. The strategies used to select translators and others members of the translation team can also be used to train them, as relevant, in the unique aspects of survey translation. The term "source language" used below refers to the language out of which a translation is made. The term "target language" is used to refer to the language into which a translation is made.

Figure 1 shows translation within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To locate potential candidates for a team translation effort and to select the most suitable from among these; to brief selected translators on general features of relevance for survey translation and on specific features of the study; and to engage and brief relevant other members of the team.

1. **Search for translators in contexts in which they are likely to work, advertise, or acquire their translation training.**

Rationale

At the selection stage it is important, whenever possible, to have multiple candidates from whom to choose. A team effort also requires more than one translator. Organizations that employ or train translators and associations with which translators register or advertise are likely places to begin locating translators for the language(s) required.
.

Procedural steps

- Identify likely organizations, associations, and places where translators register and advertise. Local options may vary greatly: search the internet and telephone directories, newspapers, and trade journals, and contact any local chambers of commerce, publishers, medical institutions, international firms, advertising companies, and places of higher education, as available, for help in making contact.

- Compose and write a job description. Post this at any place identified as potentially relevant. Send the description to any contacts made in organizations. If appropriate, include in the advertisement a request for help in locating suitable people.

- Utilize your own organizational and personal networks. Post the advertisement or job description within your own institution, and ask people you know to suggest contacts.

Lessons learned

- In some locations it may be difficult to find trained translators, either in general or for a language you require. In this case, proficient bilinguals may be the only personnel available. Follow through with them as many of the selection and briefing steps as possible.

2. Require candidates to submit application materials prior to the job interview.

Rationale

Information about a candidate's experience and training and specimens of previous translation work help decide whether a candidate merits consideration. If there are numerous applicants, these materials can be the basis for selecting people to interview.

Procedural steps

- Identify the application materials required in the advertisement. If contact is not made though an advertisement, provide candidates with the job description and request delivery of the application materials before arranging an interview.

- Ask applicants to provide the following:
  - An outline of their training and experience in translation for the languages involved (source and target). This should include the kind of translations the applicant has worked on.
  - Specimens of any recent work if possible.
  - Recent references relevant to the job application.
  - Details of their computer skills and access to computer technology.
  - Details of their work experience.
  - Details of their education in general.
  - Details of how, when, and where they acquired competence in the source and target languages.

Lessons learned

- Application materials only tell part of the story; avoid hiring on the basis of these alone. Translations delivered for inspection are, for example, not produced under team translation conditions, nor can you know precisely who contributed to their production.

- It is important to identify whether candidates are currently working in the source and target languages, or whether their exposure and use of one or the other lies in the past. Translators should ideally be embedded in the target culture and language, as well as fully conversant with the source language and, as relevant, the culture from which it springs. It is also important to ensure that applicants are competent in both speaking and writing the target and source languages.

- Although language competence in the source and target languages does not guarantee that someone can translate, it is a prerequisite. If bilingual individuals without translation training represent the highest level of expertise available in a given context, select from these, using the materials described in Guidelines 4 and 5, and train them intensively.

- Avoid engaging someone simply on the basis of recommendations whenever possible. If there are people with whom, for whatever reasons, the project team is expected to work, evaluate these people to ascertain their skills and possible language expertise. In looking for translators, you may also find suitable candidates for back-up personnel.

3. If working with translation agencies, require reference materials and specifications for both the agency and the translators.

Rationale

The professionalism of the agency needs to be verified, as well as the suitability of translators employed for the survey project. Team translation requires the translators to be available for meetings. Make sure that any agency involved understands and accepts the requirements (see Translation).

Procedural steps

- Ask agencies to provide the following information about themselves:
  - A list of clients and contact options.
  - A list of projects (the agency experience record).
  - References from recent representative clients.
  - Years of operation.
  - Information about the business focus and personnel in the agency (for example, whether the owner or manager has a translation background and whether translation is a central part of the agency's activities).
  - Any agency sub-contracting procedures relevant for your project.
  - The agency's procedures for hiring and training translators.
    - How they find and select translators.
    - How they train, if they do so.
    - How they monitor translation performance (who monitors, and how).
  - How they intend to accommodate the team translation requirements of your project (meetings, repeated access to the same translators, etc.).

- Ask agencies to provide the translator materials outlined in Guideline 2 in preparation for the selection interview(s).

Lessons learned

- The cost differential between translators working as self-employed professionals and those provided by agencies greatly depends on the individual context. The same holds with regard to quality. In general, agencies pay translators less than what independent translators working full time earn. Competent translators may nonetheless work with agencies. Agencies, for example, can provide a steady flow of work.

- Agencies initially reluctant to cooperate on requirements for team translation may later develop into valuable and reliable partners.

4. **Select translators on the basis of submitted materials and their performance in the interview.**

Rationale

The interview is the opportunity to explore and verify information provided in the application and to test performance in tasks needed for a team translation effort.

Procedural steps

- Appoint one or more people with expertise in survey translation and the languages in question to conduct the interview (typically, senior translation reviewers).

- Organize the interview in such a way that candidates actually demonstrate their competence on the spot, including their ability to produce translations, review existing translations, and accept critiquing of their translations, as well as indicate their knowledge of relevant tools, etc.

- Use the following indicators as the basis of evaluation criteria for selecting any given translator:
  - Current knowledge of and competence in the source and target languages and cultures.
  - Translation and review performance on test materials.
  - Experience and expertise in translation.
  - Knowledge of translation tools.
  - Team suitability.
  - Computer skills and access to computer technology. This may be a requirement in many projects.
  - Knowledge of and experience with translating surveys.
  - Availability and salary/payment requirements.

Lessons learned

- Extensive translation experience in one very specialized field may be a drawback for working on survey translations. Someone with years of experience in legal translation may be unused to the everyday language and tone often aimed for in survey translation.

- Experience in producing survey translations should not be taken as proof of suitability, as many survey translations are poor.

- Without briefing, translators unfamiliar with surveys may not recognize key measurement features. At the interview, assessment should focus on the demonstrated ability to understand the source text and render it fluently in the target language, as well as the ability to identify problems for translation or adaptation and to ask relevant questions. Training on survey measurement features can follow, if a candidate is hired.

- It is important to try to assess whether a candidate seems likely to work successfully as a member of a team.

5. **Brief translators on general features of surveys relevant for survey translation, as well as on specific features of the given study.**

Rationale

Briefing translators helps them to read, understand, and translate questionnaires as instruments of measurement. Translators need to be able to recognize the design features and various components of surveys in order to handle them appropriately. For example, survey questions have special vocabulary and syntactical features that may run counter to normal written language; instruments have sections addressed to different audiences (interviewer, respondent, programmer, etc.); and questions and answer scales reflect measurement goals that an untrained reader might not perceive for what they are.

Procedural steps

- Use specially developed materials or real questionnaires in source and target languages to brief translators on the following:
  - Different components of a questionnaire.
    - Questions, instructions, explanations, answer scales, fills, annotations, sections for official use, programmer instructions, formatting conventions, house-style requirements, etc.
    - Vocabulary requirements for the target population.
    - Level of vocabulary, as well as regional vocabulary considerations (see Language Harmonization).
  - Explain the notion of questionnaire modes and details of the mode for the project at hand (e.g., oral or written presentation, branching presentation of answer options, web-based response features, etc.).
  - Answer scale designs and their purposes.
  - Surveyspeak, that is, the special features of questionnaire language as found in source and target language questionnaires.
  - Adaptation and any feedback procedures to be followed.
  - Translation documentation and the procedures to be followed.

- The notions of response styles and social desirability, as well as any feedback required from translators in these situations.
- The purpose and procedures of any pretesting planned.

Lessons learned

- Careful briefing is important to guide translators' perception of questionnaires and ensure consideration of both respondent needs and questionnaire designers' needs in translations.

- Without briefing, translators will translate according to the text models and text types with which they are already familiar. Unless they are reminded that the instrument is intended for oral presentation, for example, they may produce one more suited for processing as a written text.

- Briefings should include motivating information to encourage translator commitment and care. Survey translation may call on translators to work repeatedly on the same questions; this iterative process may run counter to their expectations. If they are informed about the high-stakes nature of a survey and the survey costs involved should questions go wrong, they understand repetitive aspects of team procedures better.

6. Identify and engage suitable other personnel required for the translation effort: the senior reviewer—who may also coordinate the project—the adjudicator, and substantive experts. Translation and Translation Appendix A outline the tasks and procedures involved.

Rationale

Finding good translators is only one requirement to produce suitable target language instruments. The other personnel should be chosen with care so as to bring together the skills and knowledge required for the project, as outlined in Translation.

Procedural steps

- Identify the procedures to be undertaken and the skills required for this as described in Translation and Translation Appendix A and seek suitable personnel.

- Require these personnel, as appropriate, to demonstrate their abilities for the tasks in which they will be engaged, possibly along the model outlined above for translators.

- Tailor their briefing and training to the contributions they will make. Ensure this includes a general overview of the planned translation project phases, procedures, and responsibilities.

- If there are people with whom, for whatever reasons, the project team is required to work, meet with and evaluate these people to ascertain their skills and possible language expertise.

- Increase the size of the team as necessary to ensure the right mix of skills is available. Not everyone will be required at all times throughout the project (see Translation).

Lessons learned

- The senior reviewer and the translators are likely to be the people most important for translation quality; it makes sense to select the best people available.

- Training and briefing can greatly improve the performance of individuals and the team.

7. Use documentation as a deliberate quality assurance and control tool to enhance selection, training, and briefing and to record performance.

Rationale

Selection is partly based on reviewing documentation submitted on team members' performance and experience. It is also partly based on candidates' performance on materials and documents presented at selection and training meetings. Thus selection materials serve multiple functions. First, they allow selection committee members to prepare for the selection process, permit comparisons of candidate experience and performance, and are the basis of benchmarking. Later, selection materials used to test ability and understanding can function as training and briefing documents.

Procedural steps

Previous guidelines indicated the kinds of material to request of candidates and what to prepare for selection, testing, and briefing.

Lessons learned

- Over time, an array of materials can be assembled. Documents produced for one round of selection and briefing can be used again for other projects.

- Materials from surveys can be good resources.

- For some translation performance testing or briefing, it may be easier to create examples and tests.

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).
Adjudicator	The person who signs-off on a finalized version of a questionnaire (see Adjudication).
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Mode	Method of data collection.

Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Response styles	Consistent and stable tendencies in response behavior which are not explainable by question content or presentation. These are considered to be a source of biased reporting.
Reviewer	Person who participates in the review of translations in order to produce a final version (see Appendix A of Translation).

Social desirability bias	A tendency for respondents to overreport desirable attributes or attitudes and underreport undesirable attributes or attitudes.
Source language	The language in which a questionnaire is available from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Surveyspeak	The special features of survey language (lower pronominal anaphor, for example) as found in source and target language questionnaires.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Team translation	Team approaches to survey translation and translation assessment bring together a group of people with different talents and functions in the team so as to ensure the mix of skills and discipline expertise needed to produce an optimal translation version in the survey context. Each stage of the team translation process builds on the foregoing steps and uses the documentation required for the previous step to inform the next. In addition, each phase of translation engages the appropriate personnel for that particular activity and provides them with relevant tools for the work at hand.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Translator	The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other tasks such as reviewing and copyediting.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.

References

[1] Harkness, J. A. (1996, August). *Thinking aloud about survey translation*. Paper presented at the International Sociological Association RC 33 Conference on Social Science Methodology, Colchester, UK.

[2] Yang, Y., Harkness, J. A., Ching, T-Y., & Villar, A. (2010). Response styles and culture. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 203–223). Hoboken, NJ: John Wiley & Sons.

VIII. Translation: Translation Management and Budgeting

Janet Harkness, Dorothée Behr and An Lui

Introduction

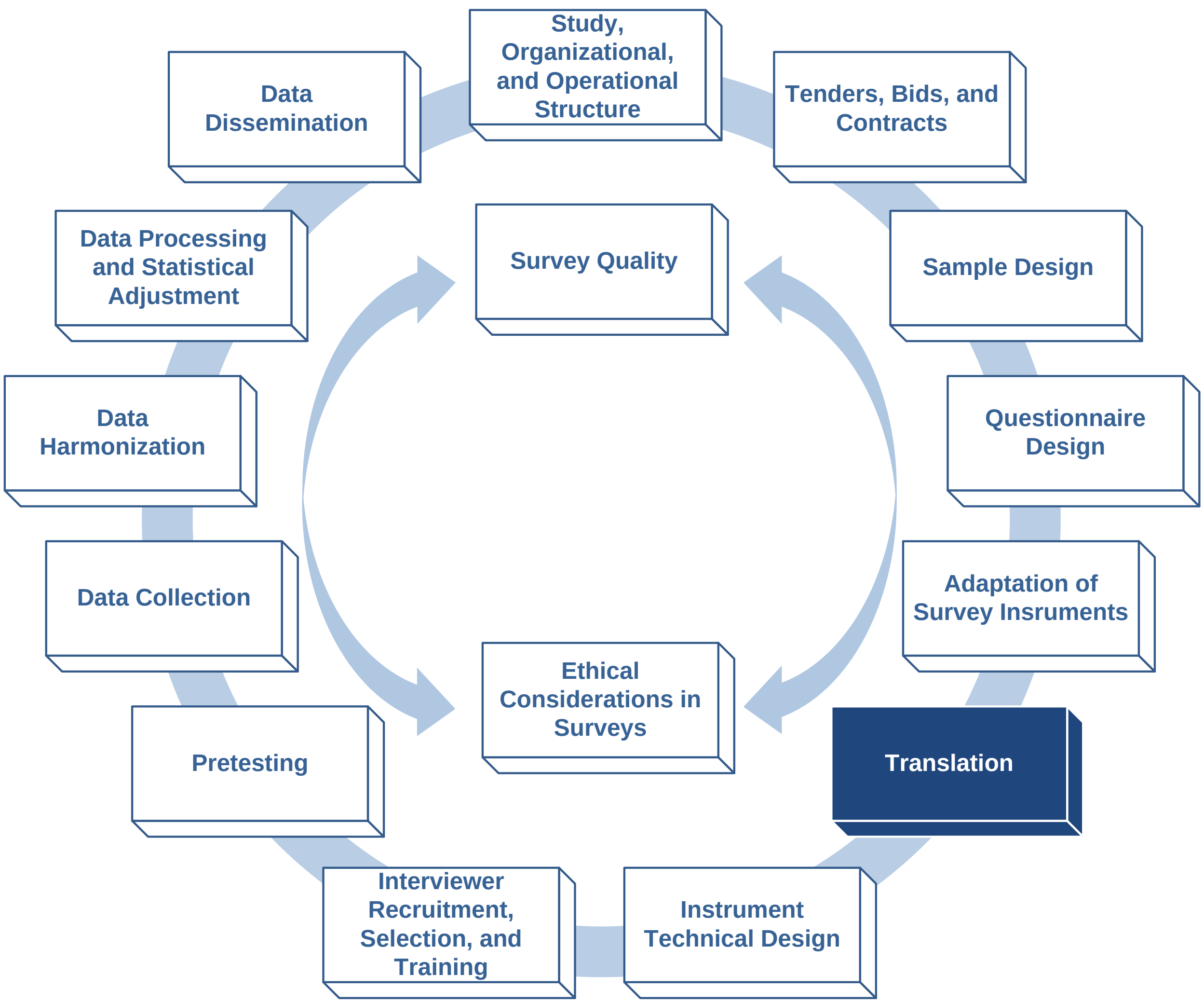
The section describes models of budgeting resources as well as budget items that may need to be included for translation (see Tenders, Bids, and Contracts for overall survey budgeting).

There is no one costing 'recipe' for all projects. The organization and scope of the translation project will determine the structure and complexity of the budget planning. For example, in a centrally organized and centrally financed project, management may be asked to specify what funding resources are needed for top-down pre-specified procedures. Alternatively, a project at local level may be asked to organize, conduct, and budget for one or multiple translations. Depending on how various levels of the project are organized, their local level costing may be needed to estimate required costs for just one translation or be used by a central national team organizing and budgeting for a number of translations for within-country fielding. Alternatively, such costs may be needed by an international team coordinating and budgeting for a multi-country project.

In order to be of relevance for projects of various sizes and budgets, the guidelines here do not assume sophisticated project management tools for translation development. They do, however, refer to the potential of such and other options (see Translation Tools). Large-scale projects on very tight timelines are likely to have such tools.

Figure 1 shows translation within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

1. Determine the project management form and the required personnel.

Rationale

Project management may vary according to the organization and scope of the translation project. In large translation efforts, centrally organized studies, and in translation projects conducted by a large organization, a coordinator may be appointed to manage the translation effort of all the languages. Additional coordinators may manage individual languages. When translation is organized at the national level and only involves the language(s) of the country, preexisting staff may take on the function of project manager.

Procedural steps

- Identify the management required or specified.

- Identify or appoint one or more project manager(s) as appropriate.
  - If several people are involved in managing the project, ensure, if possible, that one person has ultimate responsibility for signing-off on decisions, meeting deadlines, delivering products, etc.
  - Keep clear records so someone else can take over if this proves necessary.
  - If several people share the work and responsibility, set up clear sharing, delivery, and checking procedures. This reduces the likelihood of omissions and oversights when work is passed back and forth.

- Identify costs for such personnel as well as management components, such as communication, offices, and meetings.

- Identify any overhead costs not already covered.

- Explore the potential and limitations of management systems such as described in Translation Tools and determine whether any such system will be used.

- Budget for organizing and undertaking all relevant steps above.

Lessons learned

- The level of detail involved in translation project management can be easily underestimated. Good management tools are important; they need not necessarily be sophisticated technology.

- Large-scale projects will benefit from content management tools such as described in Translation Tools.

- Large-scale projects will benefit if the development of translations can be integrated into a system also managing the development of any source questionnaire, as described in Translation Tools.

2. Identify the material for translation and the language(s) required.

Rationale

The nature and the scope of the material determine which translation procedure to adopt, the number and kind of key players involved, and the schedules and budgets required.

Procedural steps

- Identify the material that must be translated. Apart from the questionnaire itself, translations may be needed of interviewer manuals, contact forms, information leaflets, and programming instructions. Some may call for a combination of local adaptation and translation.

- Establish how many languages are involved and identify any special requirements, such as interpreters for unwritten languages and word lists for interviewers working in regional dialects.

- Identify any material already translated which will be considered for re-use; assess the quality of this material and its suitability for re-use in some form.

- Select translation procedures on the basis of the material required and other relevant project considerations (see Translation and Guideline 3).

- Determine whether special tools or software are to be used in the translation development process and whether these involve costs for the project (see Guideline 6 and Translation Tools).

- Decide how translation costs are to be calculated (see Appendix A).

- Budget for preparing materials for the translation process and any preparatory steps, such as creating templates or inputting source text to software.

Lessons learned

- Some materials requiring translation can be easily forgotten. For example, if each country programs its own computer application, the programming instructions will require translation. Underestimation results in underbudgeting, not just of costs but of personnel and time.

- Questionnaires often have repetitive elements. If these can be identified ahead of time, consistency can be improved and, often, costs reduced. Payment for handling repetitive elements should also be determined (see Appendix A).

- It is important to identify clearly any sections which are not to be translated for both the budget staff and the translators.

- Shared languages which are to be harmonized will call for different budgeting. Draft translations in such instances may be cheaper but additional procedures for harmonization may increase costs again, depending on the procedures followed (see Language Harmonization).

- Good planning and preparation of material to be translated and good briefing and training are investments which can reduce later costs and improve the quality of the translation. However, such preparation must also be included in the budget.

3. **Identify the translation procedures to be followed and the human resources needed and budget accordingly.**

Rationale

The translation protocol chosen impacts the number and kind of people involved and time allocations required, as well as management, meeting, and communication costs. Translation procedures may be prescribed or selected according to the nature of the material to be translated. Low priority material might be produced by just one translator.

Procedural steps

- Determine what procedures will be followed for translating the identified materials.

- Determine what people need to be involved. Plan for translation, review, and adjudication, copyediting, formatting and, if appropriate, the programming of computer applications (see Translation).

- Identify personnel already available and any that need to be recruited for the translation project.

Lessons learned

- Different procedures may be required by different organizations and project specifications. Large educational testing projects, such as TIMSS, typically include a review and revision component undertaken by a commercial company. The World Health Organization (WHO) WMHI project required a harmonization meeting for Spanish versions. For some of its instruments, the Gallup Organization hires a commercial company to organize translators and translations, while Gallup personnel closely monitor the output. The Survey on Health, Ageing, and Retirement in Europe requires participating countries to use a common translation tool [1]. Each of these factors can affect meetings, training, the preparation required, and the degree of external versus internal outlay called for, as well as the number and kind of people involved in activities.

- The intensive, and possibly more costly, procedures chosen for one set of materials may not be needed for all the materials.

4. Determine the scope of selection and briefing meetings.

Rationale

Careful translator team selection and briefing is essential. Meetings for these purposes should be included in the budget (see Finding, Selecting and Briefing Translation Team Members).

Procedural steps

- Unless you are working within a framework that provides both the materials for selection and briefing and the protocols for these steps, budget for planning and developing these materials and protocols.

- Include outlay for selection and briefing meetings in the budget.

- Include any advertising and networking costs involved in this.

- Decide whether or not in-house training is required.
    - This will depend upon the study needs and the qualifications of the translators and any other personnel involved.

Lessons learned

- There are few selection and briefing resources publicly available for survey research translation. These can be developed from existing surveys.

- Physical meetings may be costly; training-the-trainer meetings may be of questionable suitability. Webcasting and webinars require advance preparation and time zone scheduling but may be one viable option for a worldwide project.

- Regional meetings (in whatever form) may prove more effective than too-large meetings across a project. In this case, it would be useful if at least one experienced person were able to be involved in all of the regional meetings.

5. Determine the nature and scope of review/adjudication meetings.

Rationale

Review and adjudication discussions are central to the quality of the final translation product and should be included in the budget.

Procedural steps

- Identify the number of meetings required, the form of the meetings, and the people who must be involved.
- Consider any catering, travel, or accommodation costs related to physical meetings and any other costs related to virtual meetings.
- Develop a time schedule and plan for the meetings.
- Determine the time and resources required to plan, conduct, and report on the meetings.
- Reserve funds for planned meetings after the main translation phases (e.g., after pretesting), as well as for unexpected meetings to resolve last-minute problems.

Lessons learned

- If personnel charges different rates at different times, meetings that need to take place during evenings or weekends may be more costly.
- Time-zoning may also need to be considered.
- Working days, public holidays, and “office hours” differ across countries.
- See Language Harmonization for details on this and an indication of what it could mean for budgeting.

6. Budget for materials that may need to be acquired for the project.

Rationale

Any special resources, such as software, language aids, or digital recorders should be budgeted for.

Procedural steps

- Determine whether or not materials such as the following are needed and already available:
  - Dictionaries.
  - Manuals for translator training and briefing.
  - Software or licenses (translation tools, project management tools, webcasting).
  - Notebooks or computers.
  - Projectors.
  - Digital recorders.

- If they (or other materials) are not available but will be needed, budget accordingly.

Lessons learned

- It may be difficult for a coordinator to identify or acquire materials with which he or she is not familiar and is uncertain how to locate.

- It is a good idea to check that technical components and equipment are compatible with existing equipment at intended locations before purchase. It is also useful to check that any equipment purchased has a reasonable shelf-life.

Appendix A

Estimating translation costs

Translation costs can be estimated in a number of ways. Basic ways to calculate translation costs used in various fields of work are outlined in Table 1 below, which also indicates which approaches could be useful for calculating translation costs in survey research.

Payment basis	**Explanation**	**Limitations**	**Relevance for survey research**
Number of words translated	Such calculations are normally based on the number of words in the source language. Independent of translator, this can differ considerably from the number resulting in the target language.	In questionnaires, segments of the source text are often repeated (e.g., answer scales, instructions, question stem introductions). Paying for translation of each repeated word/segment occurrence after word/segment occurrence would inflate translation costs.	Payment based on the words translated could be used in surveys for draft translations for team review. However, decisions are necessary if texts are highly repetitive about how reoccurring segments or repeated technical terms should be treated as regards payment. Since these segments/components require processing by translators, they should be paid for in some fashion. These segments should usually cost less than translated text that is not repeated. This needs to be negotiated. Extra activities required (e.g., documentation; adaptation suggestions) need to be factored in to rates paid.
Number of pages translated	An average page (whatever that is) is taken as the basis of calculation.	The amount of text on one page varies for multiple and obvious reasons. This is not a recommended	It is quite unsuitable for surveys and the multiple modes and formats survey instruments involve.

		basis for estimation.	
Number of lines translated	A line is defined as having certain properties (number of characters or number of words) and the defined line is assigned a price. Translations are estimated accordingly. It is possibly easier for some clients to envisage lines than word counts, although familiarity with word processing software can be expected to alter that.	Definitions are needed for what counts as a line and what is under a line. Payment by line requires the text to be organized in lines with the length of a line defined for purposes of payment. Material not organized in lines requiring translation must be calculated differently.	It is quite unsuitable for surveys. Survey instruments would not count well in terms of lines, even if these were defined.
Time spent on translation	An hourly going rate is determined.	Slow translators will cost more than fast translators. Speed may be a factor of experience. One translator cannot reasonably be expected to produce good work 11 hours a day and those hiring on an hourly basis should recognize this and monitor for it.	Team translations that require translators to attend meetings must calculate at least some of their costs per hour. Briefing and training sessions should be paid at an hourly rate. For an experienced and trustworthy team, an hourly rate is often the best solution. Benchmarking across colleagues upon occasion, however, (timed translation) could be a useful and salutatory performance check.
In surveys: the number of questions in an instrument or the number a	This is not a standarizeable unit to base costing calculations on.	One question may be short and easy to translate, another consist of multiple parts with	Not recommended. If clients require some idea of costs, a sample of various lengths of questions

respondent is typically presented with		technical terminology (health or insurance questions). Questions may include instructions, explanations and interviewer guidance that are also translated. All questions and all related material require translation.	could be translated and calculated in terms of time taken. This could provide a client with an indication of costs for a survey in the "question number" framework he/she has. On this basis, an hourly rate might ultimately be calculated.
Estimated duration of a interview/ application		In as much as a 10-question no filter survey is likely to take less time to translate and to field than a 50-question no filter survey, there is a weak relationship between the number of questions in a questionnaires and the extent of the work involved to translate the instrument. However, questions may include instructions, explanations and interviewer guidance that are also translated. All questions and all related material require translation irrespective of what filters mean respondents usually answer.	Not recommended for survey work.

Apart from the extent of work to be translated, numerous factors affect what a translation will cost. Table 2 outlines additional factors relevant for estimating costs for survey translations.

Table 2. Factors affecting translation costing

Factor	Comment
Availability of translators for the languages involved	It is easier in given locations to find good translators for some language pairs than for others. The more difficult it is in a location to find someone for the language pair, the more expensive the payment expectations may be. The costs for translations for English into Korean or Vietnamese, for example, are likely to vary depending on where translators are sought. Some language pairs may be expensive in almost every location. It could always be e difficult to find translators for a translation from Hungarian into Maltese, for example, or certainly more difficult than a translation from English into Spanish. Hungarian and Maltese are spoken by relatively small numbers and the likelihood of finding good translators diminishes accordingly.
Local standards of pay	These can vary greatly around the world. Some organizations aim for the same going rate (however decided) for all locations; the promise of a steady flow of work to translators might help an organization implement this model. Other organizations and projects try to optimize across locations, paying more in one location than in another and adjusting their decided going rate (however determined) on the basis of local rates of payment and local expectations.
A need to accommodate regional variants of a language	If a project needs to capture suitability for multiple regional variants of a language (Spanish, French, or German, for example), this will require more translators or language advisors to be involved than would otherwise be the case. Harmonization meetings and their outputs (see Language Harmonization) may need such translator input, even if not always in person.
Difficulty of text type	Conventionally some text types (specialized fields with special jargon) can command a higher rate of pay than do more everyday or accessible text types. Even if the rate were the same, more difficult texts could take longer and increase costs in that way. Benchmarks of difficulty are usually related to specialized vocabulary and possible specialized constructions. In surveys, the quality of source questions, target population needs and cultural distance from that assumed by the source questionnaire, and variation in questionnaire complexity are examples of factors which can add to difficulty. However, in terms of vocabulary and sentence

	structure, many questionnaires would not be considered to be difficult texts. What makes questionnaires difficult to translate is less the complexity of language used than the measurement goals pursued.
Translation mode	Oral forms of translation (on sight oral and interpreted) may command higher rates of pay than do written texts.
Experience of translators and others involved	Experience may impact speed of translation and deliberations, as well as the quality of decisions. This will affect total time needed.
Payment decided for any repeated text segments	If a survey instrument has many repeated sections (e.g., question introductions always framed similarly, frequent repetition of answer scales), this should be calculated in to reduce costs.
Time available for the translation	Express delivery or "rush jobs" normally cost more than does work submitted so as to allow the translator to fit it into his/her normal schedule.
Additional services required beyond translation	Translators can serve multiple functions beyond producing translations, either subsequent or parallel to translation. Apart from involvement in a team translation procedure (see Translation), proofreading, copyediting, and questionnaire formatting in the translated language are all tasks translators are sometimes asked to undertake. These would add to the payments made to translators, possibly also booked as "translation costs".
Training and briefing on special features of the translation	Time needed for this will be add to the final costs but improve quality and perhaps speed of the translation process.
Any software expenses	Software or license purchases may also be booked as part of the translation budget

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Consistency	Consistency is achieved when the same term or phrase is used throughout a translation to refer to an object or an entity referred to with one term or phrase in the source text. In many cases, consistency is most important with regard to technical terminology or to standard repeated components of a questionnaire. Reference to "showcard" in a source questionnaire should be consistently translated, for example. The translation of instructions which are repeated in the source text should also be repeated (and not varied) in the target text.
Content management	The software and procedures used to capture, save, organize, and distribute information in digitalized form.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.

Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Team translation	Team approaches to survey translation and translation assessment bring together a group of people with different talents and functions in the team so as to ensure the mix of skills and discipline expertise needed to produce an optimal translation version in the survey context. Each stage of the team translation process builds on the foregoing steps and uses the documentation required for the previous step to inform the next. In addition, each phase of translation engages the appropriate personnel for that particular activity and provides them with relevant tools for the work at hand.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Translator	The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other tasks such as reviewing and copyediting.

Unwritten language An unwritten language is one which does not have a standard written form used by the native speakers of the language.

Variance A measure of how much a statistic varies around its mean over all conceptual trials.

Word list When regional varieties of a language are to be accommodated, a word list can be created of the words that are required for specific varieties of a language. They can also be incorporated into computer applications of an instrument. A word list can be a useful resource for interviewers. They cannot, however, address challenges faced when regional varieties differ in more radical and structural ways from one another.

Further Reading

Cantor, S., Byrd, T., Groff, J., Reyes, Y., Tortolero-Luna, G. & Mullen, P. (2005). The language translation process in survey research: A cost analysis. *Hispanic Journal of Behavioral Sciences, 27*(3), 364-370.

Matis, N. (2006). *Translation project management*. Retrieved April 26, 2010, from http://www.translation-project-management.com/translation-project-management

Rico Pérez, C. (2002). Translation and project management. *Translation Journal, 6*(4), Retrieved April 26, 2010, from http://accurapid.com/journal/22project.htm

VIII. Translation: Translation Scheduling

Janet Harkness and Dorothée Behr

Introduction

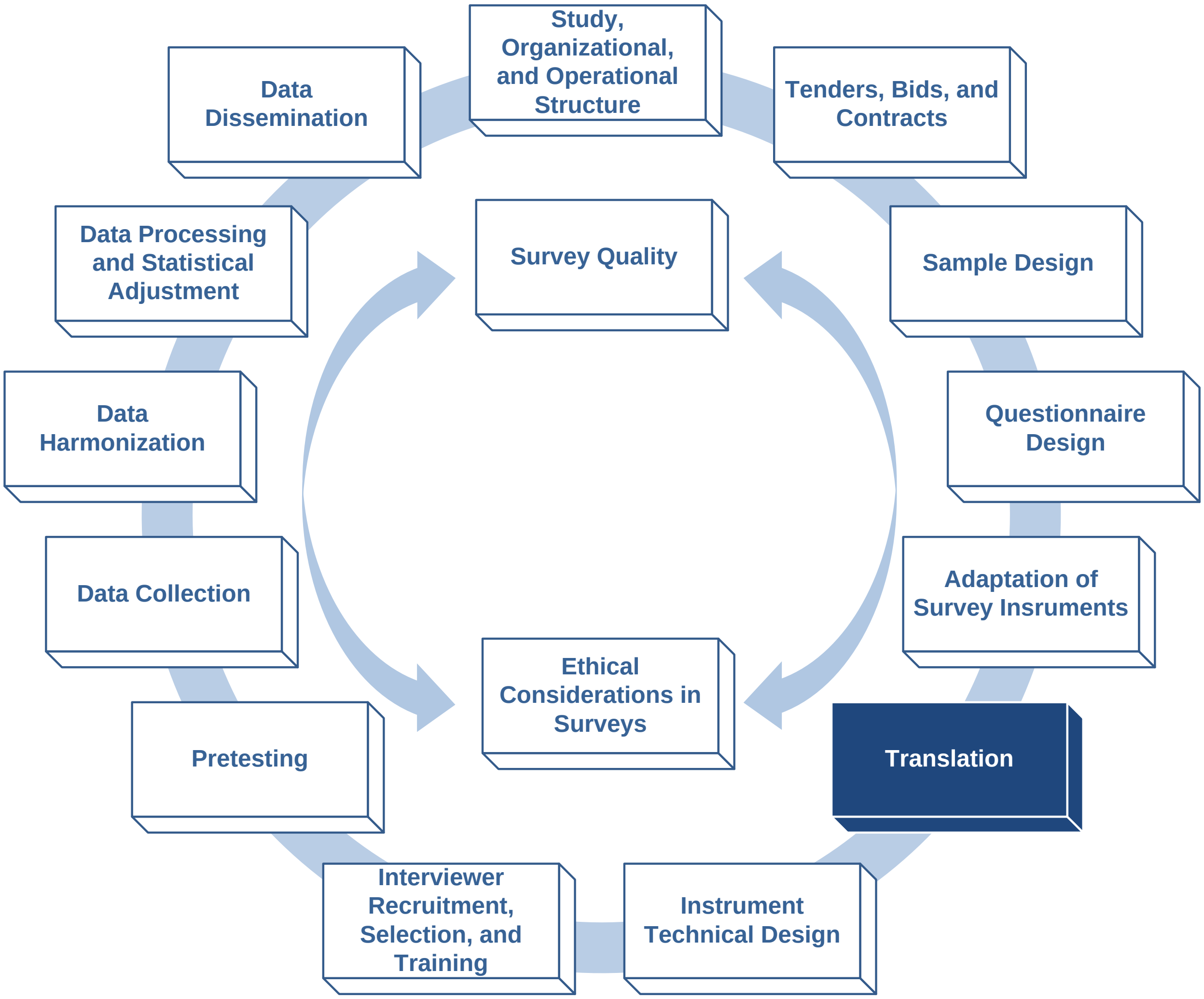
This section discusses scheduling the translation effort. Scheduling in a multinational project very much depends on how the translations figure in the project as a whole. They might, for example, be anticipated in features of the questionnaire design. There may be centrally approved specifications for how they should be conducted and when; and there might be centrally organized quality monitoring procedures. When translations are produced centrally for a multinational project, it is likely that a document management system is used in the production and scheduling of source questionnaires and translations.

The following guidelines focus on translation efforts managed at the local or national level. This will be the normal procedure for many projects. However, many of the points considered would also need to be addressed in projects using centralized development and management systems. When translation is carried out at the local level, differences and deviances across local schedules will affect timing and milestones for the general project.

No units of time per task are identified here since time required depends upon the length, the repetitive nature, and the difficulty of the material to be translated, as well as on the number and experience of those involved.

Figure 1 shows translation within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

1. **If possible, schedule translation after the source questionnaire has been finalized.**

   Rationale

   The exact length, nature, and coverage of the source instrument cannot be known until the instrument is finalized. All of these affect planning, scheduling, and quality procedures. Consistency lists, glossaries, and checklists, for example, are harder to implement without a finalized version of the source instrument. Material still to be determined may affect existing parts of the questionnaire and implementing adjustments later is complex and error-prone. Organizing translation procedures is also more complicated with regard to split options, language harmonization, and iterative review. These challenges are greatly increased if the instrument in question is long and has many submodules.

Procedural steps

- Make the importance of having a finished source version clear to those involved in procedures that impact its completion and aim to schedule accordingly.

- Optimize scheduling of the source questionnaire to accommodate translation as relevant and possible.

- Optimize scheduling of all steps related to translation.

Lessons learned

- Many steps can be completed before translation begins. Provided the nature and scope of the material is clear and the languages required can be specified, translation team members can be selected and briefed and some tools prepared.

- Time constraints may require translation to start with only a pre-finalized source text or with parts of the source text still missing. In such cases, mechanisms should be in place to efficiently and consistently update the source text and to inform all team members of the changes (see Translation Tools).

- A first round of translation can be followed later with a second round. This increases costs but can resolve problems arising from working on partially finished instruments.

2. **If possible, schedule translation when the source questionnaire, although complete and "finalized", can still be adjusted if problems are encountered.**

Rationale

Careful question design and pretesting can help identify problems in the source questionnaire. Nonetheless, some problems often become apparent only when translating into multiple languages. If adjustment can still be made to the source questionnaire and integrated in the translated questionnaires, quality and comparability can be enhanced.

Procedural steps

- Schedule sufficient time between finalizing the source questionnaire and fielding in any location to permit feedback on the source questionnaires resulting from translation.

- Optimize scheduling of the source questionnaire.

- Optimize scheduling of all steps related to translation.

- Identify how and to whom feedback (i.e., information about perceived difficulties) is to be conveyed.

- Establish schedule deadlines for feedback.

- Emphasize that timely feedback is essential.

Lessons learned

- Since problems related to the source instrument may only become apparent when translation begins, researchers sometimes recommend advance translation [1] [2].

3. **Schedule time to find, select, and brief translation team members.**

Rationale

Source text quality and client specifications impact the potential quality of translations. Apart from these, however, translation quality depends to a large extent on the competence of the staff involved. It is important to allow sufficient time to recruit and select the best possible people for the job.

Procedural steps

- Consult the guidelines in Finding, Selecting and Briefing Translation Team Members and set the time frame appropriately.

- Include time for material preparation for these procedures (see Finding, Selecting and Briefing Translator Team Members).

Lessons learned

- Finding, selecting, and briefing the translation team can be done before the source text is finalized, provided the language(s) and the nature of the instruments to be translated are sufficiently known.

- Engaging people already familiar with translation team procedures may reduce time for some of these steps.

- Contacting translators who worked well on other kinds of projects might reduce the time involved in locating potential staff.

- It may be necessary to retrain long-established translators or other team members if the current project has different needs than those of previous projects.

4. Schedule time to prepare the translation documents.

Rationale

Essential preparation steps for the translation effort must be included in scheduling.

Procedural steps

- Prepare translation and documentation tools for translators as soon as the source text is finalized (see Translation Tools). Easy-to-use translation and documentation tools speed up the translation process and make subsequent steps more efficient.

- Prepare instructions on how to comply with and use the documentation tools.

Lessons learned

- Allow sufficient time if the tools have to be produced manually. If mistakes are made in producing templates to be used in every location, for example, later attempts to correct these across locations may be unsuccessful.

- Some preparatory work can begin before the source material is finished even if its completion has to wait on the source material.

- If tools required for the project are provided by a central coordinating center, the delivery date of these tools often determines when the translation project can start at the national or local level.

- Local teams may wish to begin translation as soon as they have the source instrument. If tools are not available when that happens, they may translate without the tools. Intended quality assurance and control steps related to tools may then not be in place.

5. Schedule time to prepare the translation instructions and assemble reference materials.

Rationale

Clear project instructions and comprehensive reference materials help translation teams to produce translations that meet the needs of the project. Preparation time and delivery dates for these need to be scheduled.

Procedural steps

- Include time to compile documentation for the team on such relevant aspects of the survey as:
  - The target population (educational background, age, vocabulary requirements, etc.).
  - The mode or modes planned and how these impact the formulation and structure of the instrument.
  - How to "read" the source materials. For example, how to recognize in the source material the intended recipient for text segments (respondent, interviewer, programmer, etc.) and how to understand specific measurement features (e.g., such multiple specifications as:"Generally speaking, on an average weekday, how many times in total do you usually ….").
  - The purpose and character of source materials (e.g., interviewer manual, showcards, computer-assisted applications, explanations).
  - As applicable, style guides, quality check lists, and glossaries.
  - As applicable, reference materials, such as parallel texts, previous source text versions, available translations of the same study, and relevant background information on the research goals.

Lessons learned

- If translation team members are poorly informed about the needs of the project, quality suffers and review and adjudication become longer and more burdensome.

- Release all materials at one time rather than sending bits and pieces to the translator teams. This makes it less likely that important materials are overlooked or forgotten.

- If some or all instructions are provided by a central coordinating center, local coordinators only need to write or assemble the materials needed at their level.

6. Schedule time to produce the draft translations.

Rationale

Quality concerns require that a reasonable time frame be determined for draft translations.

Procedural steps

- Agree on deadlines for delivery with the translators; these include the deadline for quality control (see Translation, Guideline 2) and the review deadline.

- Instruct translators to report well in advance if a time frame or deadline cannot be met, so that project management can respond accordingly.

Lessons learned

- The time frame available for production of draft translations may be very short. Translators often work on multiple projects simultaneously. The sooner they are informed about the time schedule, the easier it is for them to organize their work loads accordingly.

7. Schedule time to prepare for and hold review meetings.

Rationale

Quality concerns require a reasonable time frame for review.

Procedural steps

- (See Translation, Guideline 5).

- Include time to
  - Prepare documents for review (e.g., merge documents).
  - Send translations to all team members involved in the review.
  - Prepare for the review meeting(s).
  - Hold the review meeting(s) and refine the translation(s).

Lessons learned

- The earlier team members are informed about the time frame (i.e., the time available between receiving review documents and the review itself), the better they can prepare. This is true even if there is little time between these steps.

- The time needed for the review meeting depends on the length and difficulty of the texts to be discussed, the experience of teams, and on successful management of time during the review (see Translation, Guideline 5 and Translation, Guideline 8).

8. Schedule time for copyediting in the target language and checking against the source language. Copyediting takes place several times.

Rationale

Copyediting text produced is an essential step in quality assurance and control.

Procedural steps

- Establish the stages at which copyediting will be undertaken and schedule accordingly.

- See Translation, Guideline 5 and Translation, Guideline 8.

Lessons learned

- Equipping copyeditors with a list of the most important features to check can streamline the process and reduce time and costs (see Translation Tools).

- The last rounds of copyediting should particularly focus on anything recently changed (following review or pretesting, for example); any programming specifications; and checking against the source questionnaire or other relevant materials, such as those repeating material in the questionnaire.

9. Include time for adjudication and its documentation.

Rationale

In the course of developing the translation, multiple versions of the instrument or given questions can be generated. In order to implement quality assurance and control steps, a decision must be made and recorded about which instrument or question version is taken as the final version for a given phase.

Procedural steps

- See Translation, Guideline 6 on adjudication. Adjudication is likely before pretesting and after discussing pretesting findings. Schedule time accordingly.

Lessons learned

- The resolution of some problems from the review may take more time than expected, especially when external informants or the source text designers themselves need to be contacted.

10. Schedule time for pretesting and discussion of pretest findings.

Rationale

Pretesting is an essential component of quality assurance and quality monitoring.

Procedural steps

- Schedule time for producing a version of the instrument and any other relevant materials adequate for pretesting and for the pretesting itself. (See Pretesting) .

Lessons learned

- When multiple steps are involved in translation development (e.g., multiple languages for one location or multiple varieties of one language calling for language harmonization), the timetable for pretesting and revision can become very tight.

11. Schedule time for producing the final translated questionnaire or application.

Rationale

Completion of the translation is not synonymous with completing a questionnaire or application ready for either pretesting or final fielding and time should be scheduled for this. Final checks may again need to be made.

Procedural steps

- This step includes formatting and producing any paper-and-pencil instruments and programming any computer-assisted instruments. If provided with adequate specifications, those with experience in these areas can provide estimates of the time needed.

- Include time for any final testing required.

Lessons learned

- Mistakes can be introduced at this phase too. Incorrect photocopying or scanning of a source questionnaire page used in preparing a translated version can result in a question being inadvertently omitted, for example. Programming errors and oversights at a late stage can also negatively affect quality.

12. Schedule time for consistency checks across documents.

Rationale

If some documents are related to other documents, it may be necessary to check for consistency across them. For example, if show cards repeat questions or answer scales from the questionnaire, consistency needs to be checked across these. The same holds for documents such as interviewer manuals.

Procedural steps

- Identify which documents are involved and which sections of these documents need to be checked.

- Schedule time accordingly.

Lessons learned

- It is important to check not only for the presence of various components in the documents which need to be consistent but to check the consistency of order and fashion in which they are presented. The order of answer scale response categories could be inadvertently reversed, for example.

13. Schedule time to translate, check, and produce any other materials needed.

Rationale

If other materials are needed, then they will need to be included in the time schedule and budget.

Procedural steps

- Schedule time to
  - Determine the nature of the other materials and for which stage of the study they are required.
  - Organize and realize their translation.

Lessons learned

- If the other material is not dependent on formulation and content in the questionnaire, translation can be scheduled whenever it is expedient to meet production requirements for this material.

- If the other material repeats or depends on many questionnaire components, it is better to wait until the questionnaire translation is finalized.

- If time constraints dictate simultaneous production of such other materials and the instrument, it is wise to schedule time for later consistency checks.

Glossary

Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).
Advance translation	A quick translation is made of a source questionnaire to try to find problems that only become apparent when translation is attempted. The insights are used to modify the source questionnaire or plan for adaptation.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Consistency	Consistency is achieved when the same term or phrase is used throughout a translation to refer to an object or an entity referred to with one term or phrase in the source text. In many cases, consistency is most important with regard to technical terminology or to standard repeated components of a questionnaire. Reference to "showcard" in a source questionnaire should be consistently translated, for example. The translation of instructions which are repeated in the source text should also be repeated (and not varied) in the target text.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Document management system	A document management system (DMS) is a computer system (or a set of computer programs) used to track and store electronic documents and/or images of paper documents. The term has some overlap with the concept of Content Management Systems. It is often viewed as a component of Enterprise Content Management Systems (ECM) and related to Digital Asset Management, Document imaging, Workflow systems and Records Management systems.

Language harmonization	Language harmonization can be understood as the procedures and result of trying to find a common version (vocabulary and/or structure) across questionnaires for different regional varieties of a "shared" language.
Mode	Method of data collection.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality checklist	A checklist for quality identifies all the steps, procedures, and controls specified to ensure required procedures have been followed and their goals met.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.

Source language	The language from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Translator	The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other tasks such as reviewing and copyediting.

References

[1] Braun, M., & Harkness, J. A. (2005). Text and context: Challenges to comparability in survey questions. In J. H. P. Hoffmeyer-Zlotnik & J. A. Harkness (Eds.), Methodological aspects in cross-national research (pp. 95–107). Mannheim, Germany: ZUMA. Retrieved April 29, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial11.pdf

[2] International Social Survey Programme (ISSP). (2005, April). *Translation work group report to the International Social Survey Programme general assembly*. Mexico City, Mexico: Harkness, J. A.

VIII. Translation: Translation Tools

Janet Harkness, Dorothée Behr, and An Lui

Introduction

This section discusses tools that support survey translation, including:

- Standard reference sources
  - Dictionaries, thesauri, and other hardcopy reference materials
  - Internet and Web-based reference materials

- Standard aids
  - Checklists
  - Listservers and newsgroups
  - Standard translator procedures, such as consistency procedures

- Templates for the translation process and translation output

- Technological support, such as translator software
  - Translation Memory (TM)
  - Terminology and Alignment tools
  - Concordances

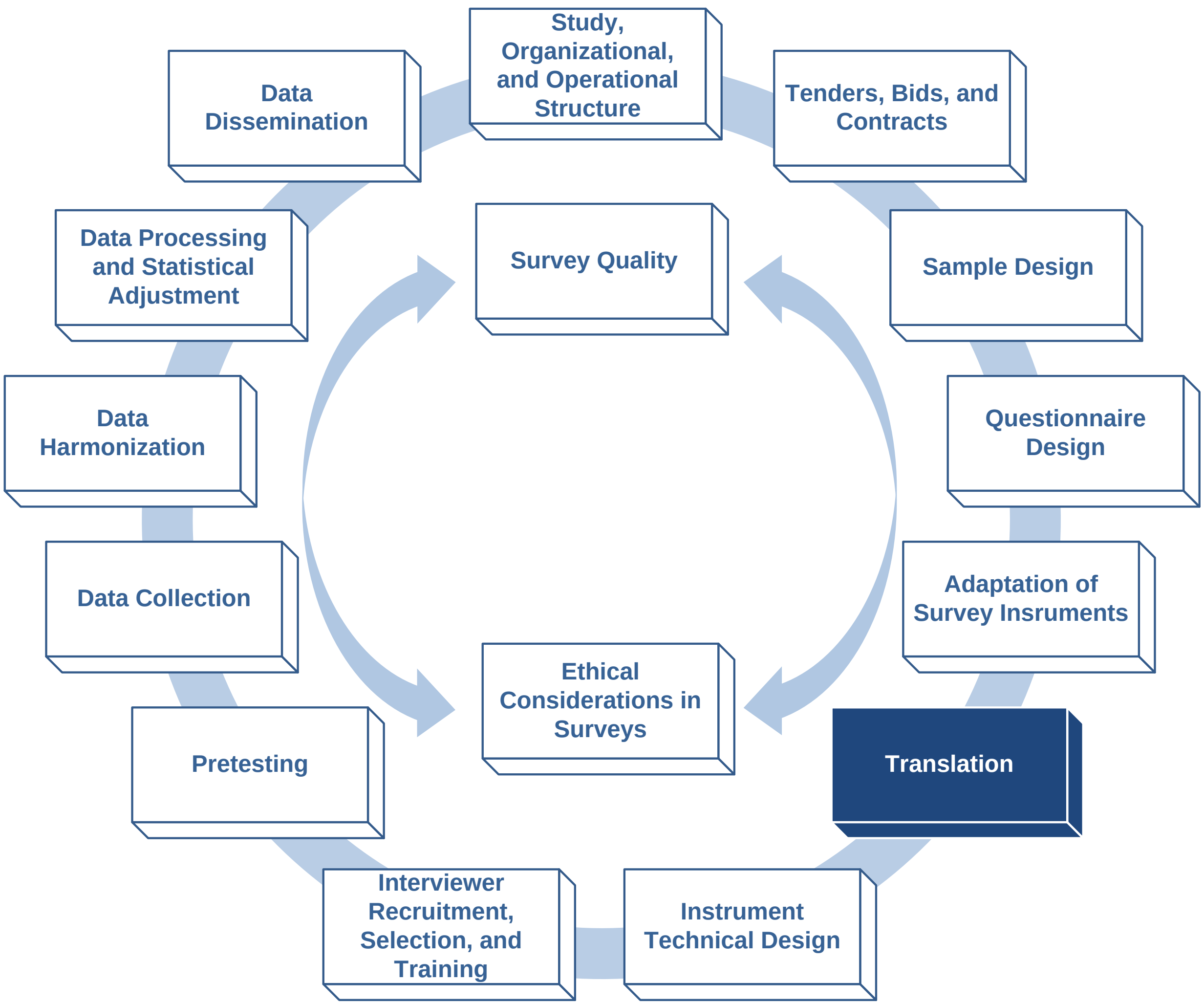
(Appendix A provides a description of various translation tools.)

Increasingly, large-scale international survey translation efforts combine source document production with that of translated versions. The source text is then entered into a content management system which anticipates the needs and documentation of later production steps in other languages [2]. In order to be more inclusive, the guidelines following do not assume such a system; they do, however, include consideration of the technological components that would be available in an integrated document production and management system [4].

Tools and aids for translation can be provided by the translation project coordinator or can be a normal part of a translator's own toolkit. Who provides what may vary by project. A project might, for example, require translators to use project-specific software to produce translations, as is the case with the Survey on Health, Ageing and Retirement in Europe (SHARE) [1].

Figure 1 shows translation within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

1. Identify relevant materials, provide them to translators, and instruct, as necessary, translators and other translation team members on their use.

Rationale

The more relevant the information and support that competent translators receive, the better they can meet the needs of a project. Other translation team members should also know about the tools and materials used in developing the translation. Depending on project organization, they will also need to use some of the tools (e.g., templates).

Procedural steps

- Consider the following materials:
  - The website (intranet and/or internet) of the survey project providing background information and documentation of the project.
  - The entire questionnaire, even if only parts of it require translation. This enables translators to:
    - See the context in which the parts to be translated belong.
    - Plan for consistency.
  - Any available sections already translated that have been vetted for quality.
    - This contributes to consistency.
    - Material not yet vetted for quality may also be provided but must be considered for re-use with great caution.
  - A bilingual glossary for any terms or phrases whose translation has already been established.
    - This helps to ensure compliance with required translations and promotes consistency.
    - It supports the review and copy-editing phases.
  - A style sheet guide, if relevant, detailing how to treat standard components of the source text (e.g., formats, use of bolding and italics).
  - Tracking documents that list major recurring elements and their location.
    - These can be produced automatically as part of a content management system and can be begun during development of the source questionnaire. Project coordinators would set the parameters for what should be included.
    - They may also be part of translation software.
    - In modestly funded projects, tracking documents can be developed manually.

- Quality checklists, created for each country's final copy editing effort. Include frequent or likely oversights in the checklist (e.g., "Check the order of answer categories").

Lessons learned

- If existing translated material that has not been vetted for quality is made available to translators, coordinators must decide whether the translators will be able to assess its quality accurately. These issues may also arise when translators access "parallel texts" (e.g., texts from other surveys) in the target language. These parallel texts might include very similar questions or include translations for standard components such as answer scales. Researchers need to be aware that existing translations may not be appropriate for their new purposes.

- The purpose of various tools and procedures called for in survey research may not be self-evident to those involved in translation production; the translation staff may need to be briefed regarding their purpose and use.

2. Provide translators and others involved in the translation with documentation tools and specifications and require them to use them.

Rationale

Documentation is part of the translation quality assurance and control framework at local and general project levels. Providing thorough documentation of decisions, problems, and adaptations at each step of the translation process guides and enhances subsequent steps. Documentation tools and specifications can ensure that each participating unit provides systematic and comparable documentation.

If the project uses a text content management system, translation documentation may be part of the development of the source document.

Procedural steps

- Clearly identify what requires translation and what does not.
  - Some work platforms allow the user to freeze sections that should not be translated.

- Produce translation templates that align source text segments, target text fields, and comments fields (see Appendix B).

- Questions, instructions, and answer scales are examples of obvious source text segments.
- Subdivisions in the template, at least to sentence level, are often useful.
- A simple MS Word or Excel table, produced manually, may suffice.
- Translation software and content management systems may produce templates automatically.

- Provide instructions for translators and any other users on how to use the templates and how to document. For example, clearly explain the kinds of information expected in any comments field.

- Hold meetings to merge template inputs. Since individual team members fill their templates, this allows them to compare options, notes, or comments (see Translation).

- Pass final output from one phase on in a modified template for the next phase of work.

Lessons learned

- The following issues apply in particular to the manual production of templates:
  - The manual production of templates is labor-intensive and calls for care. In many cases, it may be the only option. As relevant, budget for the time and effort to produce translation templates manually. Involve at least two suitable people with adequate bilingual proficiency and proofreading skills for the final proofreading effort (one reading out, the other checking).
  - Remember to check layout and format issues, not just wording.
  - Working between different source versions of a question and different translated versions within or across languages can be complicated. Any version control requires a tracking system to identify which elements should or do differ across versions.
  - Although, ideally, template production should begin after the source text is finalized, this may not always be feasible. If production of the templates starts prior to source text finalization, a tracking system for version control of templates is essential to check modifications at either the source or target text levels.
  - A procedure and protocol for alerting locations or teams to changes in either source documents or translation requirements is needed. For example, in a centrally organized project, the source text may be modified after templates have been sent out to translating locations (countries). Locations need to be able to recognize unambiguously what needs to be changed and then incorporate these changes into their templates (or at least into their

translations). The European Social Survey (ESS) alert system, as used in Round 4, for example, is accurate but not particularly user-friendly. It was produced manually and changes had to be transferred manually from source version to source version.

- Remember that copy-and-paste mistakes occur frequently. Technology (e.g., use of translation memory) may or may not make such errors more likely.

3. **Provide translators with appropriate task instructions and briefing** (see Finding, Selecting, and Briefing Translation Team Members).

Rationale

Provision of appropriate briefing and instructions helps translators and other team members understand what is required of them.

Procedural Steps

- See Finding, Selecting, and Briefing Translation Team Members.

Lessons Learned

- A hands-on presentation of activities to be undertaken or specifications to be followed is often more effective than an informational talk alone.

4. **Consider networking translation teams within the project.**

Rationale

Consultation within a language family can be helpful for all. Consultation across language families can also be of benefit, since some generic issues are shared by rather diverse languages and cultures. Although research on this is sparse, recent work suggests that a reasonably wide range of languages and cultures face similar translation challenges [3].

Procedural steps

- Decide whether collaboration is to be an official requirement or not and whether or not it must be officially documented.
  - If it is to be documented, decide on the template and detail required.
  - Official collaboration and official documentation help to unify practices across and within projects.

- Set up a protocol and schedule for sharing experiences or solutions and documenting these. Procedures described in Language Harmonization may be useful.

Lessons learned

- The publication of collaborative benefits, procedures and successful outputs experienced within one group may inspire other groups that have not have considered such collaboration. This argues strongly for documentation of work undertaken, even if it is not an official project requirement.

- Even if the languages they produce translations for differ considerably from one another, researchers may find numerous common difficulties in translating out of the source language [3].

- If researchers fielding in different regional forms of a "shared" language do not collaborate, many differences across versions may result that could otherwise have been avoided (see Language Harmonization).

5. **Make tools a deliberate part of the quality assurance and control framework for developing and checking the translated questionnaire. If possible, integrate this development with that of the source questionnaire.**

Rationale

Tools make it easier to check that procedures are implemented and facilitate checking the quality of outputs at various stages of translation production.

Procedural steps

- Determine the translation production budget and the budget available for tools of various kinds.

- Identify tools of value for the procedures to be undertaken and identify outlay for each of these. A number of these are identified in the present section; more are discussed in Appendix A.

- Obtain or create tools to be used for the translation procedures.

- Train those using the tools on their use well in advance; monitor performance as appropriate, and refresh training as needed from time to time.

Lessons learned

- Tools need not be expensive and technologically sophisticated in order to work.

- Some tools will be familiar and seen as standard aids by the translating team, while others may be unfamiliar. Good briefing and instructions will foster proper and more extensive use of tools.

- It is useful to point out the risks associated with tools as well as their advantages (e.g., “copy and paste” can be useful and can go wrong).

- Multilingual projects should investigate management systems which manage both source questionnaire development and translation development.

Appendix A

A List and Description of Translation Tools

Dictionaries: There are many kinds of dictionaries and related textbooks. Good use of dictionaries requires knowledge of their strengths and weaknesses, familiarity with the way in which dictionary entries are structured, and familiarity with the abbreviations and descriptive labels used in entries. In all instances translators ought to be familiar with the key relevant dictionaries for their area of work and know how to read and use dictionary entries.

- *Monolingual dictionaries*
  - *Source language (SL) dictionaries*
    Monolingual dictionaries list and explain the different typical meanings an SL word may have in different contexts. They may help translators check that a term fits the given context.
  - *Target language dictionaries*
    Target language dictionaries may help clarify possible meaning in the target language and provide collocations (usual word combinations). They may also offer synonyms.

- *Bilingual dictionaries*
  - *General bilingual dictionaries*
    These dictionaries list under one entry the associated terms in another language which correspond to the various meanings possible for that term. Experienced translators may use these dictionaries as checking tools or to remind themselves of definitions they may have forgotten. Inexperienced translators may mistakenly think such dictionaries can provide them with a correct word to use which they do not already know. However, if a translator does not know a word, it is dangerous for her or him to use it on the basis of having found it in a dictionary.
  - *Terminological dictionaries*
    Bilingual dictionaries can be especially useful when it comes to subject-specific terminology (e.g., medical terminology). However, languages differ in the extent to which they use technically correct terminology for subjects or prefer more everyday terms (compare “He has athlete's foot” to “He has tinea pedis”). Translators should not use terms with which they are not familiar unless they have solid evidence that these are the right terms for their needs. They may need to consult experts on a final choice. The more information a dictionary offers on the context in which suggested equivalents are embedded, the better for the translator.

- *Spelling dictionaries*
  Spelling dictionaries are useful at copyediting and proofreading stages undertaken by translators. Incorrect spelling (and punctuation, layout, etc.) can trip up both interviewers and respondents when reading questions. Incorrect spelling may also create a poor impression of the project in general. Spellcheckers included in word processors are useful but manual proofreading remains a necessary final step to recognize errors a machine cannot (e.g., form/from, on/in, healthy/wealthy)

- *Online dictionaries*
  There are numerous *online dictionaries* and thesauri. See, for example, http://www.yourdictionary.com/ or http://www.lexicool.com/.

Thesauri: Thesauri group together words of similar or related meaning. They can be helpful for finding the most appropriate word after looking up a related word known not to be quite right. The user may know the word passively and recognize it among those offered. Since a thesaurus only offers synonyms and does not define words, extensive knowledge of the language is required to identify the starting place for a search and to decide whether a term found is appropriate.

Word processors such as MS Word also offer modestly comparable functions as "Synonyms" and "Thesaurus" in at least some languages.

Internet: The Internet makes it possible to see multiple examples of words in context and to check how frequently they seem to be used. However, the Internet offers usage without quality assurance. A particular word might only appear on translated websites or on websites from countries that do not use the language in question as a first language. The word or phrase then found may not be correct for the target language or for the level of diction required for the survey.

The Internet can be used to check:

- The frequency of occurrence of particular phrases or words.
- The contexts in which words appear.
- Official terminology versus everyday terminology as evidenced by the contexts in which occurrences are found.

Listservers and newsgroups: Translators often use translation-related listservers and/or newsgroups to post questions and enquiries. Survey translation needs might not be well addressed but questions about general usage (e.g., regional terms or levels of vocabulary) could be answered. Some languages are likely to be better served than others. Sci.lang.translation is an example of a translation-related newsgroup.

Translation software: Demonstration versions of translation tools are usually available on software producer websites. Companies also usually offer to consult on prospective customers' needs. The usefulness of any of these tools for a given project depends on many factors, including the repetitive nature of the project, the scope of the project, the suitability of the tools for the specifics of a project, the budget available, and the ability of staff to work with such tools.

- *Translation memory:* A translation memory is a database that stores translations, as they are produced, for future use. "Future use" can be within the same translation, only a few minutes after first being produced or could be for an entirely new translation task months later. The source text segment and the corresponding target text segment produced as a translation are saved as a "translation unit." A segment may consist of a few words, whole sentences, or, depending on the material involved, extended stretches of text. Translation memories display source and target text segments alongside each other and thus facilitate review. In addition, they can indicate if all segments up for translation have been translated.

  When translation memory is used, it offers "100% matches" for completely identical and previously translated source text segments and "fuzzy matches" for similar, but not identical source text segments previously translated. Depending on the software used, the degree of match required in order for it to be presented to the translator can be defined. Translators accept or reject matches offered. Whatever a translator may produce as a new translation or revise by modifying an existing translation also becomes part of the dynamically created and expanding translation memory. Translations produced using translation memory can thus benefit from technology but must be driven by translator decisions. The translation memory software simply presents (offers) pre-existing translation choices for consideration. There is no quality component with regard to how appropriate the translation offered is. It is therefore essential that the memory has been created through submitting good translations.

  Properly vetted translation memories can be useful for texts that are highly repetitive and where consistency of repetitive elements is a crucial issue. They can also be of value with texts that are used repeatedly but with slight modifications.

- *Terminology tool:* A terminology tool stores multilingual terms alongside additional information on these terms, such as a definition, synonyms, and context examples. Usually, a terminology tool is used alongside a translation memory as a source of richer information.

- *Alignment tools:* Alignment tools compare a source text and its translation and match the corresponding segments. It can be used to align translations produced alongside the source text as is often the case if translation memory is not used. If desired, the aligned text could then also be imported into a translation memory and be available for future translations.

- *Translation memory versus machine translation:* Translation memories are built upon the basis of human translation. Machine translation, per se, is a fully automatized process. Quality translations never rely on machine translation alone. Survey questions are a complex text type with multiple functions and components. As a result, any reduction of human involvement in the decision-making process of survey translation is ill advised.

- *Concordance function:* This software feature allows the translator to search for terms within the translation memory: the contextual usage of a given word is then displayed, much as in a concordance.

- *Concordance:* A concordance is an alphabetical list of the words or major words used in a body of work (a “corpus”) alongside their immediate contexts (the words occurring before and after). For translators, they can clarify the contexts in which words are usually used. It is possible to buy large language corpora (collections of spoken and/or written language) and apply a concordance tool to these. However, there are no corpora available to date of survey questions and their translations that consist only of good questions. Question banks available contain questions that exist irrespective of formulation considerations. Simply collecting questions would not suffice. As is the case for translation memory, creating a useful corpus would entail checking the quality of the question formulation or the translation of any given question because the purpose of the corpus would be to indicate good formulations and appropriate contexts. However, question banks that exist in various places could be a place to begin work on creating a quality survey question and survey translation corpus.

- *Translation management*: In addition to facilitating translation, tools are available that facilitate project management. Most of the commercial packages listed in Further Reading offer such management tools.

Appendix B

Template 1 is typical of templates used in the ESS in rounds 1–4 for draft translations. The source questionnaire has been entered in the template in distinct sections. Each translator enters his/her translation in the template and provides commentary. For later stages in the translation process, similar templates retain information from each foregoing stage and add columns for outcomes and comments on the current step (see Template 2).

Template 1: Extract from a translation template from the ESS Round 4 for one draft translation (core module B)

	Source English Section B	**Routing**	**Draft Translation 1**	**Comments**
B above B1	**Now we want to ask a few questions about politics and government**			
B1	How interested would you say you are in politics – are you…			
I in B1	**READ OUT…**			
RC	very interested,			
	quite interested,			
	hardly interested,			
	or, not at all interested?			
	(Don't know)			
B2	**CARD 6**			
	How often does politics seem so complicated that you can't really understand what is going on?			
	Please use this card.			
RC	Never			
	Seldom			
	Occasionally			
	Regularly			
	Frequently			
	(Don't know)			

B = Bridge; CI = Coding / Design Instruction; I = Interviewer Instruction; RC = Response Category; RI = Respondent Instruction

Template 2 illustrates possible headings for a template bringing together two draft translations for a review meeting based on Template 1.

Template 2: Headings and columns required for a team review meeting

	Source English Section B	**Routing**	**Draft Translation 1**	**Comments**	**Draft Translation 2**	**Comments**	**Review version**	**Comments from review meeting**
B above B1	**Now we want to ask a few questions about politics and government**							
B1	How interested would you say you are in politics – are you…							
I in B1	**READ OUT…**							
RC	very interested,							
	quite interested,							
	hardly interested,							
	or, not at all interested?							
	(Don't know)							
B2	**CARD 6**							
	How often does politics seem so complicated that you can't really understand what is going on?							
	Please use this card.							
RC	Never							
	Seldom							

	Source English Section B	Routing	Draft Translation 1	Comments	Draft Translation 2	Comments	Review version	Comments from review meeting
	Occasionally							
	Regularly							
	Frequently							
	(Don't know)							

B = Bridge; CI = Coding / Design Instruction; I = Interviewer Instruction; RC = Response Category; RI = Respondent Instruction

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Bilingual glossary	A glossary is a list of words or phrases used in a particular field alongside their definitions. Glossaries are often found at the back of a specialist or academic book as an appendix to the text. A bilingual glossary lists special terms used in a particular field in two languages. A key notion or concept present in one language for a given field may not have a ready single match in a given other language.
Consistency	Consistency is achieved when the same term or phrase is used throughout a translation to refer to an object or an entity referred to with one term or phrase in the source text. In many cases, consistency is most important with regard to technical terminology or to standard repeated components of a questionnaire. Reference to "showcard" in a source questionnaire should be consistently translated, for example. The translation of instructions which are repeated in the source text should also be repeated (and not varied) in the target text.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Content management	The software and procedures used to capture, save, organize, and distribute information in digitalized form.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.

Document management system	A document management system (DMS) is a computer system (or a set of computer programs) used to track and store electronic documents and/or images of paper documents. The term has some overlap with the concept of Content Management Systems. It is often viewed as a component of Enterprise Content Management Systems (ECM) and related to Digital Asset Management, Document imaging, Workflow systems and Records Management systems.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality checklist	A checklist for quality identifies all the steps, procedures, and controls specified to ensure required procedures have been followed and their goals met.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.

Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Source document	The original document from which other (target) documents are translated or adapted as necessary.
Source instrument	The original instrument from which other (target) instruments are translated or adapted as necessary.
Source language	The language in which a questionnaire is available from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the precise estimation and reduction of the mean square error (MSE) of statistics of interest.
Translator	The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other tasks such as reviewing and copyediting.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.

References

[1] Amin, A., & Balster, E. (2010, March). *Managing translations in a single CAPI implementation across SHARE countries.* Paper presented at the International Workshop on Comparative Survey Design and Implementation, Lausanne, Switzerland.

[2] Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.

[3] Harkness, J. A., Chin, T-Z., Yang, Y., Bilgen, I., Villar, A., Zeng, W., et al. (2007, March). *Developing answer scales in multilingual studies*. Paper presented at the Fifth International Workshop on Comparative Survey Design and Implementation, Chicago, IL.

[4] Harkness, J. A., Dinkelmann, K., Pennell, B-E., & Mohler, P. Ph. (2007). Final report and translation tool specifications (Deliverable No. 5), *European Social Survey Infrastructure (ESSi) project: Improving social measurement in Europe*, Mannheim, Germany: ZUMA. Retrieved April 30, 2010, from http://www.csdiworkshop.org/pdf/2008TranslationBlueprint/2008transblue.pdf

Further Reading

Austermühl, F. (2001). *Electronic tools for translators.* Manchester: St. Jerome.

Kussmaul, P. (1995). Text analysis and the use of dictionaries. In P. Kussmaul (Ed.), *Training the Translator* (pp. 105–126). Amsterdam: Benjamins.

Mossop, B. (2007). *Revising and editing for translators*. Manchester: St. Jerome.

Ruthven-Stuart, P (2000). Concordances. Retrieved May 03, 2010, from http://www.nsknet.or.jp/~peterr-s/concordancing

SDL Trados. (2010). *Translation memory*. Retrieved April 26, 2010, from http://www.translationzone.com/en/Products/translation-memory/default.asp

For examples of CAT (computer-aided translation) tools:

- *Across*: http://www.across.net/en/index.html
- *Déjà Vu*: http://www.atril.com/
- *MetaTexis*: http://www.metatexis.com/
- *MultiTrans*: http://www.multicorpora.ca/
- *SDL Trados*: http://www.sdl.com/en/
- *Transit:* http://www.star-group.net/ENU/translation/translation.html
- *Wordfast*: http://www.wordfast.net/

IX. Instrument Technical Design

Sue Ellen Hansen and Hyun Jung Lee

Introduction

One may view technical design and implementation of a given survey instrument separately from questionnaire design per se (see Questionnaire Design). Instrument technical design focuses less on questionnaire content and much more on the design of the actual survey instrument that delivers the questionnaire content. In this sense, technical design includes the format, layout, and other visual aspects of the presentation or context of survey questions, such as how prior and following questions appear related or not related to some other questions. In some instances, questionnaire design and technical design overlap. Mode decisions, for example, may shape the technical format of questions as well as their wording.

These guidelines will use the more general terms “survey instrument” or “instrument” when describing procedures or features that apply to technical design of both paper and computerized instruments, and the term “application” — which suggests the need for at least some programming — when discussing procedures for development of computerized instruments. When there is a need to distinguish between types of computerized instruments, such as computer-assisted (computerized, but not accessed via the Internet) and Web instruments, reference will be made to the mode-specific type of computerized survey.

Study design decisions related to mode have an impact on instrument technical design requirements (see Data Collection). Such decisions include whether the survey is self-administered or interviewer-administered and whether it is administered on paper or computerized. If the survey is self-administered, a decision must be made about whether it should be a paper (by mail) or a computerized survey. For a computerized survey, whether it is a computer-assisted self-interviewing (CASI) instrument or a Web instrument may affect programming costs and the computer user interface—that is, what respondents see on the computer screen and how the computer interacts with them.

If the survey is interviewer-administered, decisions may have to be made about whether the instrument should be computerized or paper and whether it should be in person or by telephone, and there may be special techncial design considerations associated with each of those decisions, as discussed below.

Study design also involves decisions about data output, coding, and data documentation (see Data Processing and Statistical Adjustment and Data Dissemination). Thus, design decisions may have an impact on technical instrument design, which affects survey implementation primarily in three ways:

1. How easy it is for an interviewer or a respondent to use the survey instrument and to provide appropriate responses (the “usability” of the instrument, which can help minimize user burden).
2. How easy it is to program a computerized instrument and to test it.
3. How easy it is to code, output, analyze, and document survey data.

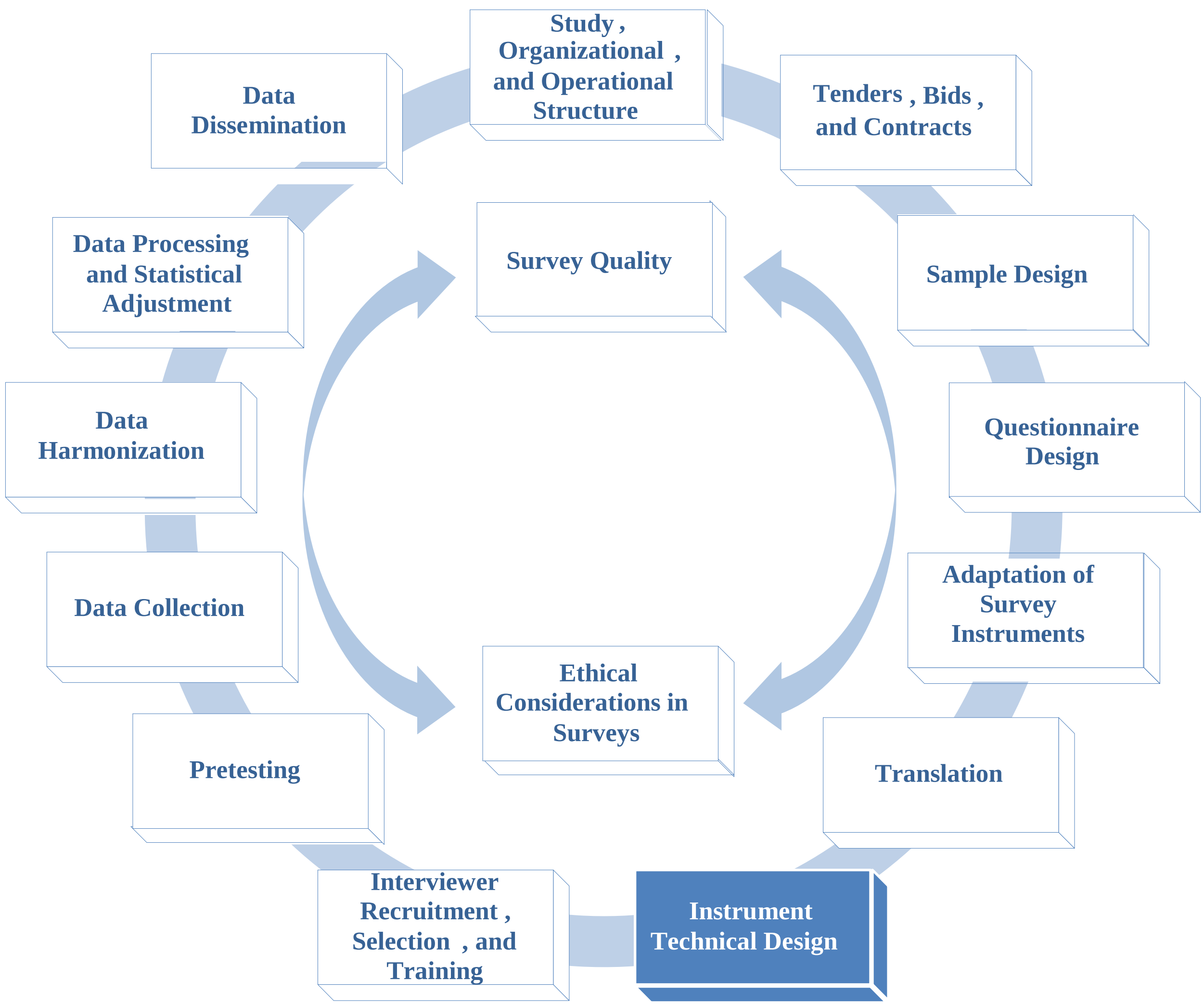
An instrument’s technical design may either lead to or minimize measurement error, including error resulting from cognitive processing, context effects, and interviewer effects. In the case of cross-cultural survey research, problems in each of the different technical implementations of survey instruments may lead to different errors. For instance, local implementations could increase inerviewer or respondent burden that will lead to cognitive processing errors or even terminated interviews. Poor design of survey instruments may also increase nonresponse error at the levels of the household or respondent (unit nonresponse) or the survey question (item nonresponse).

These guidelines are intended to help cross-cultural research coordinating centers and individual survey organizations understand instrument technical design requirements for cross-cultural surveys and how to approach creating instrument technical design specifications, whether at the centralized or local level, or both. Study design may dictate how much is specified at the central level and how much is left to local survey organizations. While there may be flexibility in this regard, it is important that technical design across local surveys leads to survey data that can be compared across cultures. For example, question labels should be consistent across survey implementations. Differences across cultures may lead to adaptations in technical design across surveys. In such cases, it is important to document the reasons for adaptation.

Figure 1 shows instrument technical design within the survey production process lifecycle (survey lifecycle)as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

.

Figure 1. The Survey Lifecycle



Guidelines

Goal: To minimize measurement error, nonresponse error, and respondent and interviewer burden due to technical instrument design, and thus maximize the amount of valid and reliable information obtained within an alloted budget and time and at the specified level of precision.

1. **Ensure that technical instrument design is appropriate to the method of administration and the target population.**

   Rationale

   Design requirements for self-administered surveys differ from design requirements for interviewer-administered surveys. Self-administered surveys have no interviewer to help repair misunderstandings. There is also limited opportunity to “train” respondents on how to respond to a self-administered survey. Computerized instruments, which involve human-computer interaction, call for design features that facilitate such interaction. Target population characteristics (education, survey experience, literacy, computer literacy, etc.) influence instrument design decisions. For example, self-administered surveys are useful only if administered to populations with high literacy rates; computerized surveys require target populations with familiarity with computers, or situations in which data collection can be facilitated by interviewers, teachers, or other aides.

Procedural steps

- Determine whether to develop an interviewer- or self-administered instrument and whether to use a paper or computerized instrument. Some points to consider are:
  - Self-administration may lead to better data quality for surveys with extremely sensitive questions, such as drug abuse or sexually deviant behavior [27].
  - Self-administered components can be combined with interviewer-assisted components of surveys.
  - An interviewer-administered instrument would be better when there is a need to explain concepts and probe responses.
  - Paper instruments may be less costly to develop, but entail additional data entry costs after data collection, and may affect the timeliness of data dissemination (see Data Dissemination).
  - There can be infrastructural constraints in some contexts that make it difficult to collect data with telephone or Web survey instruments (e.g., the lack of high telephone or Internet penetration) (see Data Collection).
  - Some countries or regions may not have the professional expertise in place to do computerized surveys.
  - Computer-assisted and Web instruments require programming, but Web surveys generally are less costly, and don’t necessarily require professional programmers for basic programming. On the other hand, if not programmed well, they may introduce higher costs during data processing.
  - Interviewer administered computerized instruments may lead to higher data quality in long and complex surveys or those with embedded experiments (for example, randomizing the order of questions or response options).
  - Computer-assisted self-interview (CASI) and Web instruments should be shorter and less complex in order to minimize respondent burden, but still allow for embedded experiments.

- Determine the appropriate instrument design for the method of administration (see Data Collection):
  - It is important that interviewer-administered instruments make it easy to perform required tasks in the order in which they are expected to be performed. For example, interviewer tasks such as referring to show cards or other aids, reading questions, providing definitions, probing responses, and recording responses should be displayed in the order of their likely occurrence. This is true in both paper and computer-assisted instruments.
  - Similarly, it is important that self-adminstered instruments make it easy for respondents to recognize instructions (such as “Select one”), and to read questions, navigate correctly through the instrument, and enter responses [12] [13]. For example, instructions should appear where they are needed, such as “Start here” before the first question, and response entry instructions after the question text (e.g., “Tick all that apply”). In addition, filter questions and instructions to skip questions should be avoided in paper self-administered instruments because they can lead to response errors.
  - Whether interviewer- or self-administered, the instrument technical design should help to minimize the burden placed on interviewers and respondents, which increases as instruments increase in length and complexity.

- Determine whether there are additional design considerations related to characteristics of members of the target population, such as children, men, the elderly, or the visually or hearing impaired [11].

- Ensure that all such considerations are reflected in the technical specifications for the survey instrument (see Guideline 2).

Lessons learned

- The use of survey computer assisted methods can help camouflage complexity and facilitate the tailoring of instruments to special populations. For example, de Leeuw, Hox, and Kef [11] describe the results from a number of Dutch surveys of special populations using computer-assisted interviewing and self-administered components, in which instrument design and administration were tailored to target population needs. For example, a simple but attractive screen layout was used to survey grade school children. In addition, students only needed to use simple keystrokes to answer questions and could stop temporarily when they felt tired. As a result, item nonresponse was reduced compared to a paper questionnaire. They concluded that well-designed computer-assisted instruments both improve the quality of data and minimize the burden experienced by respondents and interviewers.

- Study design should consider the potential measurement effects that may arise from differences in methods of survey administration. A review of paradata from the European Social Survey (ESS) and the International Social Survey Programme (ISSP) revealed some differences in results across countries between those that implemented paper self-administered surveys by mail and those that used interviewer-assisted self-administered surveys or face-to-face surveys.

2. Develop complete technical instrument design specifications for the survey instrument, specifying culture-specific guidelines as necessary.

Rationale

Technical instrument design specifications guide formatting or programming of the survey instrument or application. They ensure design consistency across culture-specific instruments (to the extent possible) and facilitate post-production data processing, harmonization, documentation, and analysis (see Data Processing and Statistical Adjustment and Data Harmonization). A coordinating center's specifications should clearly outline the source questionnaire and its content, provide rules for formatting the survey instrument, and suggest appropriate instrument design adaptation strategies for other cultures. Survey agencies may have to adapt specification rules further to adhere to local standards for design of instruments and staff training and other organizational constraints. Any such adaptations should be documented.

Note that similar guidelines are necessary for a data entry application (see Data Processing and Statistical Adjustment). Generally this guideline is relevant to formatting of elements in either paper or computerized instruments, although a few may relate to only one or the other. Guideline 4 adds guidelines that are relevant specifically to computerized applications and their interface designs and to self-administered paper instruments.

Procedural steps

- Make sure that that formatting allows for cultural differences. For example:
  - Differences in the formatting of information and areas for the recording of responses [2], including:
    - Date and time (e.g., 24-hour versus 12-hour clock).
    - Calendar, holidays, and start of week.
    - Numeric formatting (e.g., thousands, million, and billion, and decimal separators).
    - Names and addresses (e.g., last name first or second).
    - Telephone numbers (e.g., with or without local prefix).

- Currency and monetary values (e.g., placement of currency symbol and negative sign).
- Sizes and measurement (e.g., metric versus imperial units, Celsius versus Fahrenheit, clothing sizes, etc.).

- Provide rules for the consistent formatting of specific text elements, such as question text, response scales, respondent or interviewer instructions, and so on. These might include, for example [9]:
  - Display question text more prominently than response options.
  - Distinguish interviewer or respondent instructions, for example, in a smaller font of a different color, or italicized in parentheses.
  - Place text elements where and in the order they are needed based on interviewer or respondent task demands; for example, in an interviewer-adminisered instrument, a show card instruction precedes question text and a probe instruction follows it.
  - Evenly space response options in a scale, grid, or table, so that they appear of equal weight or prominence.
  - Underline question text that should be emphasized.

- Provide rules for the formatting of specific question and response types and other information, and examples for each; these may include:
  - Enumerated or fixed choice response options (e.g., 1=Female, 2=Male).
  - Tick [Check / Select] all that apply.
  - Short or fixed-length text.
  - Open-ended text.
  - Numeric responses.
    - Response entry masks (e.g., __/__/____ for dates).
  - Multi-part questions and question series; for example:
    - Day / Month / Year.
    - Address / contact information.
    - Demographics question sets.
    - Amount-per-unit (e.g., income per day / week / month / year).
  - Randomly ordered questions, response options, or sections.
  - Answer scales.
    - Fully-labeled scale.
    - Partially-labeled scale.
    - Roster or grid. Rosters are tables used to collect various information in columns about entities in rows, for example gender and age (columns) about persons in a household (rows). Grids are often used for scale ratings (columns) on a number of items (rows).
  - Text fills (variable question text); for example, question text may vary based on size of household—“you” for respondent in a single-person household, and “you and your family living here” for a household with multiple persons.

- Visual or contextual indicators that help respondents or interviewers understand where they are in a question series (for example, indicating above or beside a series of questions which household member, vehicle, or source of income they are about).
- Progress indicators (i.e., a visual indicator of where the interviewer or respondent is in the instrument as the survey progresses).
  - Progress indicators are speculated to reduce breakoffs, but added graphics associated with the use of a progress indicator increases download time [7] [8].
- Question-level help (question-by-question objectives, including definitions) in paper or computerized surveys.
- Validation or consistency checks and post-collection edits. For paper instruments, these should be noted in the instrument technical design specification for use in post processing. In computerized surveys with programmed consistency checks that occur during the survey interview, there is a distinction between a
  - Hard consistency check (interviewer or respondent cannot continue until an inconsistency is resolved), and a
  - Soft consistency check (interviewer or respondent may continue without resolving the the inconsistency).

- Add information to the instrument specifications that facilitates recording responses, the linking of survey instrument information and variables in a dataset (data dictionary), and documentation of the instrument and dataset, traditionally called a codebook (see Data Processing and Statistical Adjustment and Data Dissemination guidelines; see also Appendix C). For example, specify:
  - How questions are identified in the dataset (variable names and labels), and how response categories are numerically represented and labelled (value labels) .
  - Open question formats; consider space provided, which may need to differ across languages.
  - Pre-coded response options. If necessary, specify international standards for code numbers and classifications, such as occupation, language, country of origin, and religion (for example, specifications for the European Social Survey state that codes for respondent's language(s) are based on the ISO-639-2 code frame, but use alphanumeric codes in the dataset).
  - Code number conventions (e.g., Yes=1, No=5; Yes=1 or No=2; or No=0, Yes=1). Note that code numbers are generally not shown in self-administered questionnaires. Yes=1 and No=5 is sometimes used instead of Yes=1 and 2=No to minimize error in interviewer-administered surveys. This is because the number 5 is farther away from the number 1 than the number 2 is on a computer keyboard; thus, 2 (No) is less likely to be pressed when the interviewer means to press 1 (Yes).

- Missing data categories, such as,
  - Not applicable (does not apply to the resondent; question not asked based on prior answer).
  - Refusal (respondent refused to answer question).
  - Don't know/Can't choose.
  - No answer (interviewer or respondent did not provide response, including due to errors in computerized instrument programming).

  Note that interviewing, coding , or statistical software may constrain labels used to create survey datasets. Specifications should indicate the values required in the final datasets and in final data documentation (codebook).
- Data input formats, including scales that use metaphors (such as ladders or thermometers).
- Interviewer or respondent instructions.
  - Respondent show card instructions.
  - Routing (skip or filtering) instructions.
  - Response format or data entry instructions.
- Universe statements, that is, metadata that indicates a question or question group was asked of a specific sub-group of the survey population (e.g., "Universe [for this question]: Women aged greater than or equal to 45 years").
- Variables to construct or recode during postproduction.

- Provide rules for the use of numbers, color, graphics, images, maps, and icons.
  - Ensure that numbers used in response scales visible to respondents do not have specific implications in some cultures (e.g., "lucky" or "unlucky")
  - Ensure that colors used in instruments do not have any negative connotations in specific cultures. Color has different meaning across cultures and research has found there are cultural differences in color preferences. Any choice of colors should be validated by experts on particular cultures [2] [19] [22]. This may involve harmonization to a set of "culture-neutral" colors across instruments, or adaptation of some colors across instruments as necessary. For example,
    - Red in China means happiness while it means danger in the Western countries, as well as in Japan [22].
    - White, black, all shades of gray, all shades of blue and a light yellow are preferentially used internationally [22]. However, be aware of any association of specific colors with political groups in some countires.
  - Ensure that any maps used are drawn to scale.
  - Ensure that images are displayed using comparable typographical units across survey implementations.

    - Ensure that graphics, images, and icons convey comparable meaning across cultures and do not have negative connotations in specific cultures, or adapt them as necessary.

- If using multiple data collection methods, include specifications for how question formats would differ across methods. For instance, a survey may be interviewer-administered in multiple modes (paper and computerized, or in-person and by telephone); it may be self-administered in two modes (Web and mail); or it may be self-administered in multiple modes (computer-assisted, paper, and Web). For example:
    - A computer-assisted self-interview (CASI) screen might have only one question and input field per screen (to minimize respondent burden), whereas an interviewer-administered computer-assisted screen might have multiple questions and multiple input fields.
    - Self-administered instruments may be developed without response codes (the respondent clicks on a response option, or clicks on a radio button, or checks a box), whereas some computer-assisted personal interview (CAPI) surveys may require numbered response options for entry of responses, if numbers are the only possible form of input.
    - Software constraints may also necessitate alternate specifications, for example, if different software were used for Web and computer-assisted telephone interviewing components.

- Based on the guidelines specified above, as well as the interface design and paper instrument guidelines that follow, prepare a survey instrument specification with all survey contents for the instrument as well as a data dictionary, which represents the contents of the survey dataset. Also specify the codebook metadata before data collection.
    - At the beginning of the instrument specifications, provide an overview of the survey instrument, including the order of core chapters and required placement of culture-specific chapters (see an example in Appendix C).

Lessons learned

- Seemingly small differences in instrument design across cross-cultural surveys can influence responses across cultures. For example, scales that are not formatted consistently, response options with misaligned check boxes, differences in the relative amount of space allowed for open responses, and differences in the physical placement of follow-up questions have been shown to lead to missing data or unusual response distributions across surveys [23]. For example, in the 1987 International Social Survey Programme (ISSP) there was a question on subjective social stratification. Respondents in nine countries were asked to rate themselves on a scale from 1 to 10 (top to bottom). In all countries respondents tended to rate themselves in the middle, and a small proportion of respondents rated themselves in the bottom. However, the Netherlands had 60% in the middle, compared to 72% to 84% in other countries, and had 37% in the bottom, compared to 6% to 24% in other countries. Dutch respondents did not have such a distinctive distribution on other social inequality measures. On examination, it was found that the Dutch translation was comparable to English, but the visual display of the scale differed (see Appendix D).

- On the other hand, cultural customs and norms may require using different graphic images, icons, colors, etc. For example in 2007, ISSP allowed countries to use different graphics for an ideal body shape question. See Appendix D for images used in the Austrian and Phillipines questionnaires.

3. Develop language-specific guidelines for the survey instrument as necessary.

Rationale

Different language features across cultures are important in designing survey instruments. Survey instrument designers should consider both languages and countries or cultures when developing language specifications, since there is no one-to-one match in languages and cultures. Some countries share the same language (e.g., English), but may have different language layout systems, and some use multiple languages in a country (e.g., Belgium and Switzerland). In addition, some countries have more than one script or system of writing (e.g., Japan). Therefore, consider any differences across survey implementations in scripts, character sets, fonts, text directions, spelling, and text expansions when developing instrument technical design specifications [2]. This is important for computerized instruments, since software may need to be configured and instruments reprogrammed to display languages in cultures for which they was not originally developed.

Procedual steps

- Provide instrument formatting specifications that facilitate the translation of languages (see Translation), specifying scripts, character sets, fonts, spacing, and so on, for target languages [1] [2] [18] [22] and the programming of computer-assisted instruments; formatting guidelines should address aspects of design such as:
  - Language- and region-specific characters sets.
    - The International Organization for Standardization (ISO) 8859 Character Set has language-specific groupings, for example, ISO 8859-1 for Western Europe and ISO 8859-2 for Central and Eastern Europe.

- Differences in languages and scripts; for example:
  - Japan has one language, but several scripts, which can be mixed.
  - China has one official language, Mandarin (Putonghua), seven major languages, and many dialects. Also, Chinese may be displayed in either Traditional or Simplified script.
- Differences in fonts that support different character sets; in general:
  - Avoid complex or ornate fonts.
  - Provide interline space to ensure clear separation between lines and to accommodate underlining.
  - Provide space to accommodate changes in line heights.
  - Provide flexibility in layout of the instrument to accommodate expansion or contraction of text during translation. For example, use a larger font and/or margins for an English instrument, if translating from English into other languages would increase the amount of space required for text in culture-specific instruments.
- Differences across languages in punctuation (e.g., the different question marks in English and Arabic, ? and ؟, respectively).
- Language- or culture-specific differences in the ways characters are sorted alphabetically, including diacritics (accent marks above or below letters, e.g., *é*), ligatures (multiple letters treated as single typographical units, e.g., æ, œ, and ß), character combinations (e.g., *ch* follows *h* in Czech), and uppercase and lowercase letters. For instance, the *Ä* sorts after *Z* in Swedish, but after *A* in German. This is important for computerized survey software that was designed for one type of culture but used in other cultures or countries that sort lists such as response options differently.

- Consider differences in text directionality and provide application design specifications that can be adapted to translated instruments with differing text directionality; the three types of text directionality are:
  - Left-to-right (Latin, Cyrillic, Greek, Thai, and Indic languages).
  - Left-to-right and vertical (Chinese, Japanese, and Korean).
  - Bi-directional (Arabic and Hebrew; characters displayed right to left, with Latin characters displayed left to right) .
  - Text directionality applies to displaying images. For example, in Arabic and Hebrew where, the text is read from right to left, images are also read from right to left [2].

Lessons learned

- In Asian countries, vertical text direction is seldom used for survey questions, but it is sometimes used for response options. In the 2006 East Asia Barometer survey, there were differences across countries in the use of vertical text. Mainland China and Taiwan used vertical text for response options, but Singapore did not. In the International Social Survey Programme in 2007, Japan and China used vertical text. When vertical text was more than one line, they were displayed from left to right in Japan, although they were displayed from right to left in mainland China (see Appendix E). These differences suggest both that design specifications need to reflect an understanding of how different Asian countries display text both vertically and horizontally, and that it would be desirable to pretest separately questions that differ across countries.

4. Develop interface design rules for computerized survey applications, and for self-administered paper instruments.

Rationale

Interface design has an effect on the respondent-computer or interviewer-computer interaction, influences user performance, and may affect data quality. Design should not only minimize respondent and interviewer burden and thus maximize usability, but should also be consistent across survey implementations. Therefore, it is important to provide clear guidelines for design of questions, error messages, and screen elements for computerized instruments (see Appendix A for an example of basic design guidelines for computer-assisted surveys). Note that similar rules are necessary for data entry applications (see Data Processing and Statistical Adjustment).

Many of the principles for interface design of computerized instruments are also relevant to paper instruments. They can just as easily address the usability of paper instruments, whether they are for interviewer-administered or self-administered surveys. In the procedural steps below, no distinction is made between computerized and paper instruments if a step would apply to both paper and computerized surveys. Where necessary, distinctions are made between computer-assisted and Web interface design.

Procedural steps

- Establish the key principles for design, which should lead to effective assessment of the quality of design (see Guideline 5). These include:
  - Consistency.
  - Visual discrimination among questions and related elements, so that interviewers and respondents quickly learn where different elements are located, and thus where to look for what type of element. For example, interviewer and respondent instructions may appear in a smaller text, a different font, and/or a color, to distinguish them from the question text.
  - Adherence to a culture's normal reading behavior for each language and script, based on issues such as text directionality (see Guideline 3).
  - Display of instructions at points appropriate to associated tasks.

  - Elimination of unnecessary information or visual display of other features that distract interviewers and respondents.

- Provide rules for the layout and formatting of question elements, including:
  - Question text, which should be the primary focus of a question, and its related information.
  - Response options, which should have instructions or visual characteristics that convey whether a single mutually-exclusive response or multiple responses are possible. For example, in computerized instruments, radio buttons convey there should be one response, and check boxes convey that there may be multiple responses, which can be reinforced by an instruction (e.g., *Select all that apply*).
  - Response input fields should convey the length of the response expected. For example:
    - An open-ended response area is as wide and and has as many lines as the expected length of response.
    - The width of an integer response area should be as many number of character lengths wide as the expected input, that is, one character length for a one-digit integer, a two-character length for a two-digit integer, etc.
  - Instructions, which should appear as expected in relation to task demands; for example, a reference to a respondent booklet or show card should appear before question text, and a probe or data entry instruction after question text.
  - In computerized instruments, the interface should facilitate accessing online help, through clear formatting of help text and design of navigational aids that facilitate opening and closing help text windows.
  - Error messages, warnings, and consistency checks in computerized instruments should clearly identify the nature of the problem, reflect actual question wording if necessary (e.g., for interviewer probes for more accurate responses), and convey how to resolve the problem (see [20] for examples and for more detailed guidelines on design of error messages).
  - Context markers (for example, instrument section labels, household member numbers, and so on).
  - Additional information may be required for Web self-administered surveys, such as contact information and graphic and/or text identification of the sponsoring organization.
  - In Web surveys, provide guidance on whether to use a paging versus a scrolling design [21] .
  - Provide rules for handling cultural differences, for example, differences in paper sizes for paper surveys. In such cases, provide guidance on pagination in order to avoid inadvertent context effects (for example, two related questions appearing together on one page in one country's survey and on separate pages in another).

- Provide examples of key question types and elements for all target languages and cultures, and for different types of administration if relevant (see Appendix A for examples of computerized questions and Appendix B for examples of paper questions).
- Provide examples of correct formatting of elements, for all question types (see Guideline 1) and all languages and cultures (see Appendix A).

Lessons learned

- There is increasing evidence that the visual design of computer-assisted and Web instruments can impact data quality [3] [6] [9] [10] [11]. For example, providing an input box or field that allows entry of 10 numbers with no guidance on input format can lead to poorer data quality than if the survey question more precisely calls for an integer of up to three digits; for example, instead of “20,” “90” or “100” in an entry field with a width of three (___), a Web survey respondent enters “40 to 50” in a field with a width of 10 (__________).

- Not providing rules for formatting questionnaires printed on different sized paper can lead to poorer comparability of data across countries. For example, in the International Social Survey Program (ISSP) one country lost the last item in a scale when copying the scale from A4 size paper (8.27" by 11.69") to letter size paper (8.5" by 11") [24].

5. Establish procedures for quality assurance of the survey instrument that ensures consistency of design, adapting evaluation methods to specific cultures as necessary.

Rationale

As discussed in Guideline 4, research shows that instrument technical design can affect data quality in compter-assisted or Web surveys, positively or negatively. This is also true of paper instruments. Thus, it is important that pretesting (see Pretesting) of comparative survey instruments include procedures for assessing the quality of the design of the survey instrument and adaptations for specific culture, languages, and modes, not just the quality of the content. This includes the evaluation of the use of color, graphics, images, maps, and icons. As indicated earlier, such evaluation procedures may require adaptation across cultures.

Procedural steps

- Identify a team with members that have expertise in evaluation of technical instrument design. Such experts may include substantive experts, survey methodologists, linguists, and usability professionals, and should include someone with an understanding of response styles across cultures.

- Provide a clear set of instrument specifications and/or a data dictionary for the instrument and culture-specific adaptations (per rules outlined in Guideline 2), which will facilitate testing and assessment of the instruments. Such documentation would include: question (variable) names and labels; question text, response option values and labels, numeric response formats and ranges, and specifications for the lengths allowed for open-ended question text; interviewer or respondent instructions; missing data values; skip instructions; and so on. It should enable during evaluation comparison of computerized or formatted paper instruments to instrument design specifications.

- Identify appropriate instrument evaluation procedures for the comparative surveys under evaluation. These may be more or less extensive based on whether survey organizations in the targeted cultures previously have used specific guidelines, instruments, and survey software. Most questionnaire pretesting tools (see Pretesting) may be used to evaluate instrument design as well as questionnaire content and data collection procedures. These include:
  - Expert review or heuristic evaluation, in which one or more experts evaluates the instrument design against a set of evaluation criteria or heuristics, for example:
    - Consistency and adherence to design guidelines.
    - Error prevention.
    - Usefulness of documentation, definitions, help, error messages, and other feedback to users.
    - Ease of navigation.
    - Ease of recognition of specific question or instrument elements and actions required.
  - Review of an instrument, data dictionary, or codebook to ensure adherence to instrument specifications for naming and labeling of variables and response options. This should include comparison across instruments or data dictionaries for all survey implementations.
  - Laboratory or on-site tests of instrument design with users or participants with similar characteristics to target interviewers or respondents. These are called usability tests when evaluating computer-based instruments, but they also may be used to evaluate paper instruments. Since culture-specific response styles affect how participants respond to questions about usability [4], every effort should be made to match tester and participant characteristics, language, and cultural background.
  - If feasible, incorporate methodological experiments on formatting, to assess whether aspects of formatting affect respondents differentially across cultures.

- Collect measures from all instrument evaluation procedures that will lead to informed decisions about question- or screen-specific or global design changes that need to be made (see Pretesting). Examples include:
  - Questionnaire length and section and item timings.
  - Audit trails for computer-assisted or Web applications, which can include item timestamps, keystrokes, mouse actions, and functions invoked. Gathering some of these requires programming that captures information directly from the respondent's computer (Heerwegh [16] provides sample programming code for capturing such paradata for Web surveys).
  - Behavior codes or event codes based on video or audio recordings that reflect problems using the survey instrument. Such methods are appropriate for both paper and computer-assisted instruments.
  - Qualitative anlayses of cognitive and usability testing.
  - Heuristic evaluation or expert review.

Lessons learned

- Research [5] [15] has shown that techniques for evaluating the effectiveness of paper materials and computer software work very well in the evaluation of the design of survey instruments. For example, usability evaluation methods (commonly used in the development of software to assess the quality of user interfaces) and traditional pretesting methods such as conventional pretests, cognitive interviews, and behavior coding can be used to identify instrument design problems as well as problems related to question content.

- Interviewer and participant interaction may need to be considered for usuability tests of cross-cultural design insturments. There is evidence that when an interviewer is from the same culture as participants, interviewers give more help, tell more about introductions, and encourage participants more frequently; and participants report more usability problems and give more suggestions than when an interviewer is from a different culture [25]. On the other hand, some research indicates that when interviewers are from cultures speaking different languages, participants explain more about their choices of design elements [28].

- Incorporating methodological experiments into cross-cultural surveys, whether for experiments on instrument design or other methodolgoical issues, can be difficult to negotiate. It involves agreement of funding agencies, the central coordinating center (if there is one), and the survey organizations involved. It also requires that clear experimental design specifications are included as part of the development of design specifications prepared for each survey organization (see Guideline 2).

6. Maintain complete documentation of source and target language or culture-specific instruments, including specification and design guidelines, and provide comprehensive summaries of the same for data dissemination and analysis.

Rationale

Comprehensive documentation of survey instruments or applications is an essential component of study documentation and comes into play at all stages of the survey lifecycle (questionnaire development, pretesting, data collection, post processing, and data dissemination and analysis). Complete and consistent rules for specifying and designing instruments is important (although not sufficient) to ensuring survey data meet the quality requirements of users (see Survey Quality). Documentation of instrument design specifications also plays a significant role in this regard. In cross-cultural surveys, it also facilitates the assessment of comparability of survey data across cultures. The rapid increase in computer-assisted data collection methods makes it increasingly possible to provide well-documented survey data. Based on study design, the study coordinating center, the survey agency, or both would be responsible for maintaining documentation related to technical instrument design.

Procedural steps

- Maintain documentation of the rules specified for technical instrument design.

- Maintain documentation of quality assessments of the survey instruments, and the outcomes of decisions made to revise the instrument design.

- Maintain specifications for the final source instruments, based on Guideline 1, Guideline 2, Guideline 3, and Guideline 4.These should include the instrument specifications and data dictionaries developed by the coordinating center and/or survey organizations.

- Maintain alternative specifications for target languages or cultures as necessary. For example, if the source instrument is computer-assisted, but it is necessary to develop a paper instrument for one or more locations, separate specifications should be developed for paper instruments.

- Maintain paper and/or electronic copies of all culture-specific instruments or adaptations of instruments, to facilitate comparison of technical design across culture-specific surveys.

- Maintain question-level metadata (question text, response options, instructions, text fills, population universes, definitions, etc.) in an electronic format to facilitate linking and comparing metadata for all survey instruments (e.g., eXtended Markup Language (XML) data files). If feasible, this should be part of a centralized documentation system that links question metadata and formatting with data codebooks for data disseminated. Some computer-assisted data collection software now makes this possible.

- Provide comprehensive documentation of survey instruments, based on all of the sources of documentation listed above.

Lessons learned

- Survey instrument design and documentation of design rules and specifications can affect the quality of data produced and disseminated, and the ability of users to effectively analyze survey data. Hert [17] conducted studies of users "interacting" with statistical data in order to understand how to better meet their needs. In one study she found that the completeness and quality of available question-level survey instrument documentation and metadata affected users' selection of variables for analysis. In particular, she found that users used a number of mechanisms for identifying appropriate variables for analysis, including what they knew about variable naming conventions, how particular questions relate to other questions, and even coding categories, if the question text did not provide enough information for selection. These findings reinforce the need for clear documentation of technical design guidelines and instrument specifications, and for these to be readily available to data users.

Appendix A

Technical Design Standards

Following are some basic standards or rules for design of interviewer-assisted computer-assisted instruments using Blaise interviewing software [26], which were based on initial research on developing guidelines for Computer-Assisted Personal Interviews (CAPI) [7]. These are included to convey the types of information to include in such standards; for example, display instructions in a smaller font of a different color than question text. Standards for cross-cultural studies should reflect the requirements for design of instruments across cultures, which could dictate choice of fonts, colors, and so on. The referenced standards included examples of basic screen types formatted according to the standards (see Figures A1 and A2 for selected question type examples).

Text Characteristics

- Display question text on a light background color (cream), in mixed case, and in 12-point Arial, black.
- Display instructions in **11-point Arial bold blue**.
- Display response categories:
  - those read out to the respondent, in 12-point Arial black.
  - those not read out to the respondent, in **11-point Arial bold blue**
- Use underline for emphasis, sparingly.
- Place optional text in (parentheses).
- Display in-text references to function keys and numbers to type in mixed case within square brackets, for example, [Enter], [1], [F12], and [Ctrl-R].

On-Screen Instructions and Other Information

- Place references to interviewer aids (e.g., an event history calendar or show card instruction) and the question text in the upper left corner of the screen, above the question text.
- Place instructions that precede the question flush left with the question;
- Use icons to distinguish special instructions:
  - 🕮 **Page 1**, for respondent booklet instruction.
  - ▤ **Calendar**, for event history calendar instruction, and.
  - ☑ **Interviewer Checkpoint**.
- "Bullet" all other interviewer instructions with an 11-point bold blue diamond (♦ **Enter [1] to continue**).
- Single space within an instruction and double space between instructions;
- Place an online help indicator (**[F1]-Help**) above the question on the right margin, for questions with "question-by-question objectives" (QxQ's);
- Indent instructions that follow the question.

- Place any context-related information below the question-level help indicator on the right margin (for example, changing person-level information as the interviewer navigates a household roster or grid).
- Display instructions in the order associated with required interviewer tasks;
- Include an actual question in explicit interviewer checkpoints, displayed in **11-point Arial bold blue**.
- Capitalize only key task-related action verbs (ASK, READ, ENTER, and PROBE), and only at the beginning of instructions.
- Keep instructions simple and concise.
    - Put long instructions or those not directly related to asking questions or entering responses into online help (question-by-question objectives).
- Conditional instructions start with the conditional phrase, not the action verb, and the action verb is not capitalized (e.g., conditional probes and data entry instructions).
- In probe instructions, place text to be read to the respondent in Arial black.
- Place references to respondent answers in quotation marks.

Examples of Formatted Questions

Figure A1. Example of Multiple Response Question

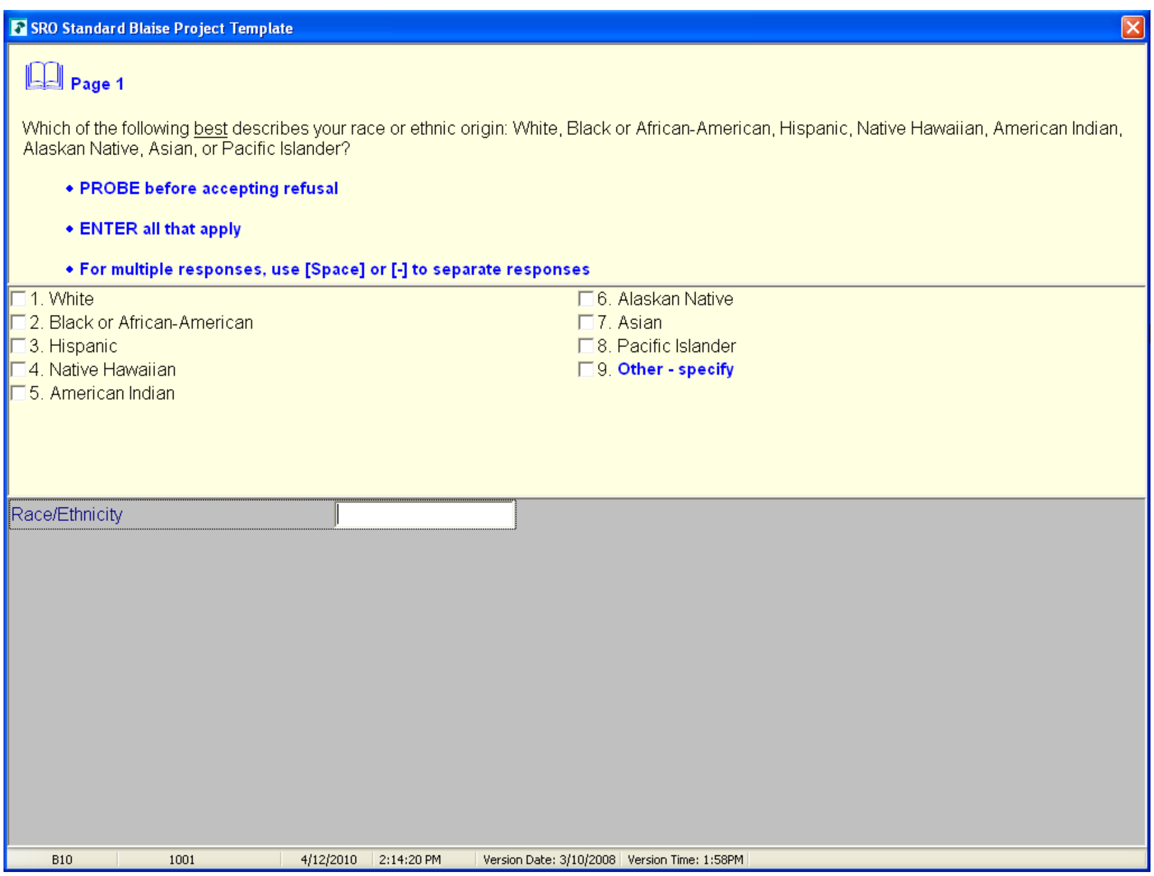
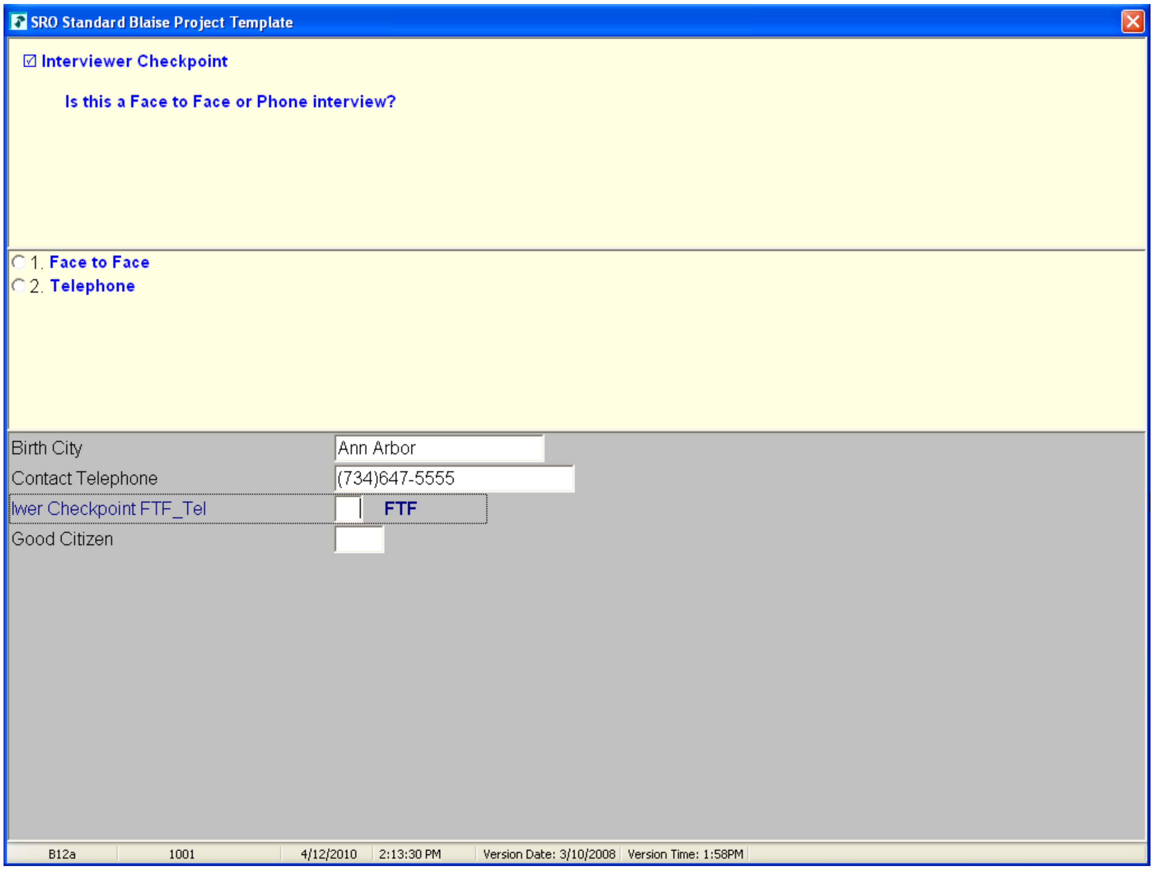


Figure A2. Example of an Interviewer Checkpoint



Appendix B

Following are examples taken from the International Social Survey Programme (ISSP) 2007 and the U.S. Census 2010 self-administered paper questionnaires. They both show instructions to the respondent, including skip instructions.

Figure B1. Example of Self-administered Questions from the Australian ISSP 2007

K4. Generally speaking, do you usually think of yourself as Labor, Liberal, National or what?

Liberal ☐
Labor (ALP) ☐
National ☐
Australian Democrat ☐
Green ☐
One Nation ☐
Family First ☐
No Party ☐ ▶ Skip to K6
Other party ☐

Please specify

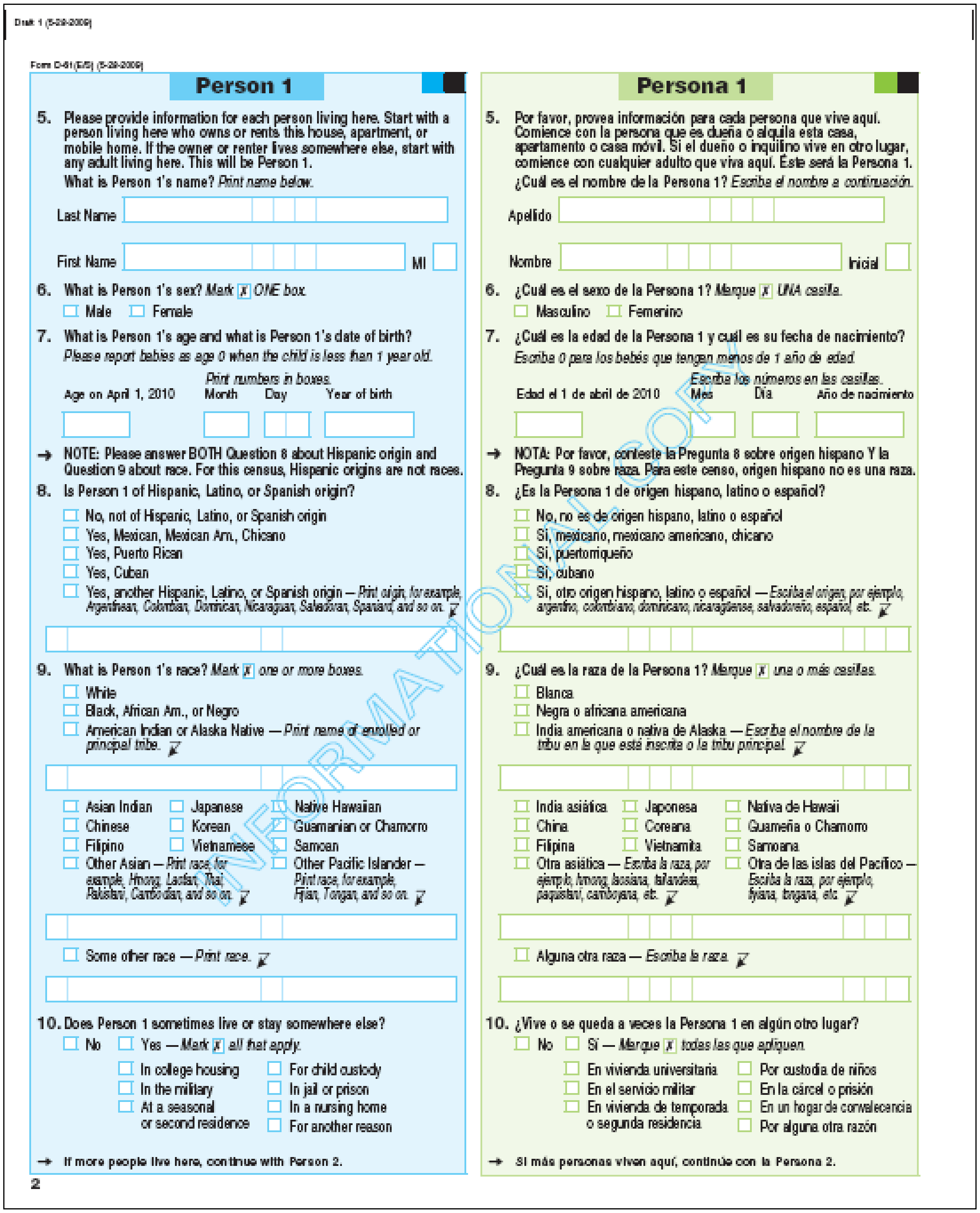
K5. Would you call yourself a very strong, fairly strong or not very strong supporter of that party?

Very strong supporter ☐
Fairly strong supporter ☐
Not very strong supporter ☐

K6. Did you vote in the Federal Election held on October 9, 2004?

Yes ☐
No ☐

Figure B2. Example of Self-administered Questions from the U.S. Census 2010 Bilingual (English and Spanish) Paper Self-Administered Questionnaire



Draft 1 (5-28-2009)

Form D-61(E/S) (5-28-2009)

Person 1

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.
What is Person 1's name? *Print name below.*

Last Name

First Name MI

6. What is Person 1's sex? *Mark* ☒ *ONE box.*
☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?
Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.
Age on April 1, 2010 | Month | Day | Year of birth

→ NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.

8. Is Person 1 of Hispanic, Latino, or Spanish origin?
☐ No, not of Hispanic, Latino, or Spanish origin
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Puerto Rican
☐ Yes, Cuban
☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark* ☒ *one or more boxes.*
☐ White
☐ Black, African Am., or Negro
☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
☐ Filipino ☐ Vietnamese ☐ Samoan
☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
☐ Some other race — *Print race.*

10. Does Person 1 sometimes live or stay somewhere else?
☐ No ☐ Yes — *Mark* ☒ *all that apply.*
☐ In college housing ☐ For child custody
☐ In the military ☐ In jail or prison
☐ At a seasonal or second residence ☐ In a nursing home
☐ For another reason

→ If more people live here, continue with Person 2.

Persona 1

5. Por favor, provea información para cada persona que vive aquí. Comience con la persona que es dueña o alquila esta casa, apartamento o casa móvil. Si el dueño o inquilino vive en otro lugar, comience con cualquier adulto que viva aquí. Éste será la Persona 1.
¿Cuál es el nombre de la Persona 1? *Escriba el nombre a continuación.*

Apellido

Nombre Inicial

6. ¿Cuál es el sexo de la Persona 1? *Marque* ☒ *UNA casilla.*
☐ Masculino ☐ Femenino

7. ¿Cuál es la edad de la Persona 1 y cuál es su fecha de nacimiento?
Escriba 0 para los bebés que tengan menos de 1 año de edad.
Escriba los números en las casillas.
Edad el 1 de abril de 2010 | Mes | Día | Año de nacimiento

→ NOTA: Por favor, conteste la Pregunta 8 sobre origen hispano Y la Pregunta 9 sobre raza. Para este censo, origen hispano no es una raza.

8. ¿Es la Persona 1 de origen hispano, latino o español?
☐ No, no es de origen hispano, latino o español
☐ Sí, mexicano, mexicano americano, chicano
☐ Sí, puertorriqueño
☐ Sí, cubano
☐ Sí, otro origen hispano, latino o español — *Escriba el origen, por ejemplo, argentino, colombiano, dominicano, nicaragüense, salvadoreño, español, etc.*

9. ¿Cuál es la raza de la Persona 1? *Marque* ☒ *una o más casillas.*
☐ Blanca
☐ Negra o africana americana
☐ India americana o nativa de Alaska — *Escriba el nombre de la tribu en la que está inscrita o la tribu principal.*
☐ India asiática ☐ Japonesa ☐ Nativa de Hawaii
☐ China ☐ Coreana ☐ Guameña o Chamorro
☐ Filipina ☐ Vietnamita ☐ Samoana
☐ Otra asiática — *Escriba la raza, por ejemplo, hmong, laosiana, tailandesa, paquistaní, camboyana, etc.*
☐ Otra de las islas del Pacífico — *Escriba la raza, por ejemplo, fiyiana, tongana, etc.*
☐ Alguna otra raza — *Escriba la raza.*

10. ¿Vive o se queda a veces la Persona 1 en algún otro lugar?
☐ No ☐ Sí — *Marque* ☒ *todas las que apliquen.*
☐ En vivienda universitaria ☐ Por custodia de niños
☐ En el servicio militar ☐ En la cárcel o prisión
☐ En vivienda de temporada o segunda residencia ☐ En un hogar de convalecencia
☐ Por alguna otra razón

→ Si más personas viven aquí, continúe con la Persona 2.

2

INFORMATIONAL COPY

Appendix C

Following are examples taken or adapted from the European Social Survey (ESS) Round 4 [29] that could be included in a coordinating center or data collection agency study rules to demonstrate instrument technical specifications for different information and question types. These can be applicable to either paper or computerized instruments.

Instrument Overview

	Q#	Topics
Core	A1 –A10	Media; social trust
Core	B1 – B40	Politics, including: political interest, efficacy, trust, electoral and other forms of participation, party allegiance, socio-political orientations
Core	C1 – C36	Subjective well-being, social exclusion; religion; perceived discrimination; national and ethnic identity
Rotating module	D1-D50	Welfare includes attitudes towards welfare provision, size of claimant groups, views on taxation, attitudes towards service delivery and likely future dependence on welfare.
Rotating module	E1-E55	Ageism covers attitudes towards and experiences of ageism, age related status, stereotypes, experience of discrimination and contact with people in other age groups.
Core	F1 – F73	Socio-demographic profile, including: household composition, sex, age, type of area, education & occupation of respondent, partner, parents, union membership, income, marital status
Supplementary	Section G	Human values scale
Supplementary	Section H	Test questions
Interviewer questionnaire	Section I	Interviewer self-completion questions

Missing Value Definitions

Not applicable: 6, 66, 666 etc.	respondent has been routed away from the question
Refusals: 7, 77, 777 etc.	respondent has explicitly refused
Don't know: 8, 88, 888 etc.	respondent has explicitly said "don't know"
No answer: 9, 99, 999 etc.	Missing data not elsewhere explained

Common Question Types

1. Interviewer checkpoints:

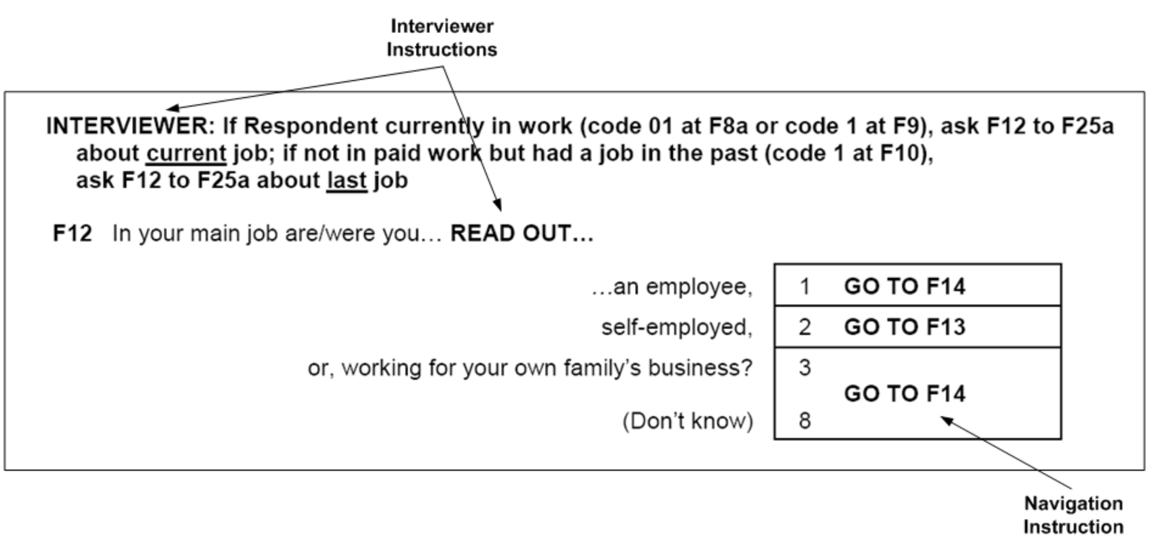
- With masks showing date format separators [□□ /□□ /□□] and response format instructions [(dd/mm/yy)]:

INTERVIEWER ENTER START DATE: □□ /□□ /□□ (dd/mm/yy)

INTERVIEWER ENTER START TIME: □□□□ (Use 24 hour clock)

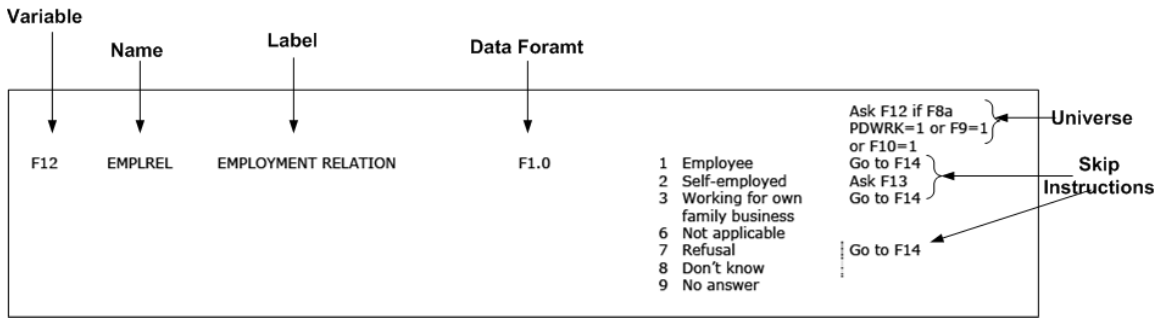
2. Closed question with enumerated response options**:**

   - Question **F12** interviewer instructions and navigation instructions (e.g., if response is **3** or **8**, **GO TO F14**).



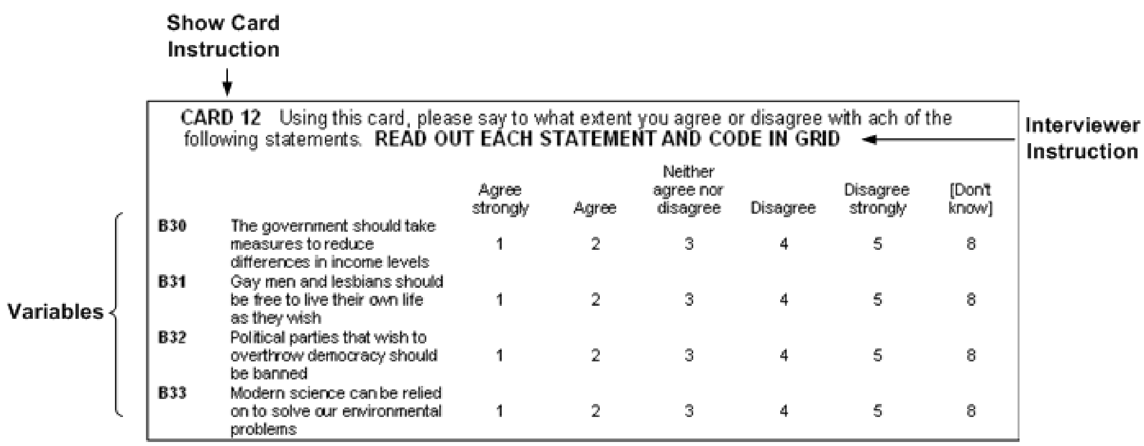
.

- Data dictionary elements for question **F12** [variable ID **F12**; variable name **EMPLREL**; variable label **EMPLOYMENT RELATION**; one-digit integer format with zero decimal places; universe (**Ask F12 if F8a PDWORK = 1 or F9=1**); response options and codes; and skip instructions]:



3. Scale Questions in Grid:

   - Questions **B30** through **B33** [show card (**CARD 12**) and interviewer instructions):



		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	[Don't know]
B30	The government should take measures to reduce differences in income levels	1	2	3	4	5	8
B31	Gay men and lesbians should be free to live their own life as they wish	1	2	3	4	5	8
B32	Political parties that wish to overthrow democracy should be banned	1	2	3	4	5	8
B33	Modern science can be relied on to solve our environmental problems	1	2	3	4	5	8

- Questions **B30** through **B33**, in the ESS Round 4 Israel Hebrew questionnaire:

[השתמש בכרטיס 12] באיזו מידה אתה מסכים או לא מסכים לכל אחד מהמשפטים הבאים:

	מאוד מסכים	מסכים	באמצע (לא 'מסכים' ולא 'לא מסכים')	לא מסכים	מאוד לא מסכים	<u>לא להקריא</u>: לא יודע	<u>לא להקריא</u> מסרב/ אין תשובה
101 **B30** הממשלה צריכה לפעול לצמצום פערי שכר	1	2	3	4	5	8	9
102 **B31** הומואים ולסביות צריכים להיות חופשיים לחיות את חייהם כרצונם.	1	2	3	4	5	8	9
103 **B32** יש להוציא אל מחוץ לחוק מפלגות פוליטיות המעוניינות בהפלת הדמוקרטיה.	1	2	3	4	5	8	9
104 **B33** ניתן לסמוך על המדע המודרני לפתרון הבעיות הסביבתיות שלנו.	1	2	3	4	5	8	9

- Show card (**CARD 12**, used for questions **B30** through **B33**):

Question(s) B30, B31, B32, B33

CARD 12

Agree strongly
Agree
Neither agree nor disagree
Disagree
Disagree strongly

- Show card (**CARD 12**) in the ESS Round 4 Israel Hebrew questionnaire:

<u>כרטיס 12</u>

מאוד מסכים	מסכים	באמצע (לא 'מסכים' ולא 'לא מסכים')	לא מסכים	מאוד לא מסכים
1	2	3	4	5

- Data dictionary (data protocol) for scale questions in grid [variables **B30** through **B33**; variable names **GINCDIF**, **FREEHMS**, **PRTYBAN**, **SCNSENV**; variable labels (e.g., **GOVERNMENT SHOULD REDUCE DIFFERENCES IN INCOME LEVELS**); single-digit integer with no decimal places; universe; response options and codes]:

						B30-B33: Ask All
B30	**GINCDIF**	**GOVERNMENT SHOULD REDUCE DIFFERENCES IN INCOME LEVELS**	**F1.0**	1	Agree strongly	B30-B33: Same format,values and Categories
				2	Agree	
B31	**FREEHIS**	**GAYS AND LESBIANS FREE TO LIVE LIFE AS THEY WISH**		3	Neither agree nor disagree	
B32	**PRTYBAN**	**BAN POLITICAL PARTIES THAT WISH OVERTHROW DEMOCRACY**		4	Disagree	
				5	Disagree strongly	
B33	**SCNSENV**	**MODERN SCIENCE CAN BE RELIED ON TO SOLVE ENVIRONMENTAL PROBLEMS**		7	Refusal	
				8	Don't know	
				9	No answer	

4. <u>Country-Specific Questions</u>

- ESS highlights country-specific questions in gray in the source questionnaire specifications, for example, variable **B12**:

B11 Some people don't vote nowadays for one reason or another.
Did you vote in the last [country] national[8] election in [month/year]?

Yes	1	**ASK B12**
No	2	
Not eligible to vote	3	**GO TO B13**
(Don't know)	8	

IF YES AT B11 (code 1)

B12 Which party did you vote for in that election?
[Country-specific (question and) codes]

Conservative	01
Labour	02
Liberal Democrat	03
Scottish National Party	04
Plaid Cymru	05
Green Party	06
Other (**WRITE IN**) ______________________________	07
(Refused)	77
(Don't know)	88

- Country-specific question **B12** in the ESS Round 4 Latvian questionnaire:

B11	**Mūsdienās cilvēki tādu vai citu iemeslu dēļ mēdz nepiedalīties vēlēšanās. Vai Jūs balsojāt pēdējās Saeimas vēlēšanās 2006. gada oktobrī?**	Jā	1
		Nē	2 ➔B13
		Nav balsstiesību	3 ➔B13
		GP	8 ➔B13
B12	*JA B11 TIKA ATBILDĒTS JĀ (B11=1)* **Par kuru partiju Jūs balsojāt šajās Saeimas vēlēšanās?**	Tautas partija	01
		Zaļo un Zemnieku savienība	02
		Jaunais laiks	03
		"Saskaņas Centrs"	04
		Latvijas Pirmās partijas un partijas "Latvijas Ceļš" vēlēšanu apvienība	05
		Apvienība "Tēvzemei un Brīvībai"/LNNK	06
		Politisko organizāciju apvienība "Par cilvēka tiesībām vienotā Latvijā"	07
		Par citu partiju (*IERAKSTIET*)	08
		Atteicās atbildēt	77
		GP	88

- Data dictionary (variables **B11** and **B12**; variable names **VOTE** and **PRTVTxx**; variable labels; one- and two-digit integer formats; response options and codes; universes; and skip instructions):

					B11: Ask all
B11	VOTE	VOTED LAST NATIONAL ELECTION	F1.0	1 Yes	Ask B12
				2 No	
				3 Not eligible to vote	Go to B13
				7 Refusal	
				8 Don't know	
				9 No answer	
					Ask B12 if B11=1
B12	PRTVTxx	PARTY VOTED FOR IN LAST NATIONAL ELECTION, [COUNTRY]	F2.0	66 Not applicable	B12: Country-specific question, see section E.1.1
				77 Refusal	
				88 Don't know	
				99 No answer	
	PRTVxx1	PARTY VOTED FOR IN LAST NATIONAL ELECTION 1, [COUNTRY]	F2.0	66 Not applicable	
				77 Refusal	
	PRTVxxN	PARTY VOTED FOR IN LAST NATIONAL ELECTION N, [COUNTRY]		88 Don't know	
				99 No answer	

Appendix D

Figures D1 through D3, taken from Dutch questionnaires of the International Social Survey Programme (ISSP), shows how different visual scales might result in distinctive response distributions. In 1987, the Dutch questionnaire used the scale that displayed a truncated pyramid, while other countries used a scale with 10 vertically stacked squares. As a result, the response distribution from the Dutch question differed from that of other countries and did not correlate well with other Dutch measures [24]. The Dutch scale for the social ladder question was later changed to more closely resemble the visual display used by other countries (Figure D3). Figures D4 and D5 show differences in graphics used for a body shape question in the ISSP 2007 Austrian and Phillipines questionnaires.

Figure D1. Social Ladder in ISSP 1987 Dutch Questionnaire

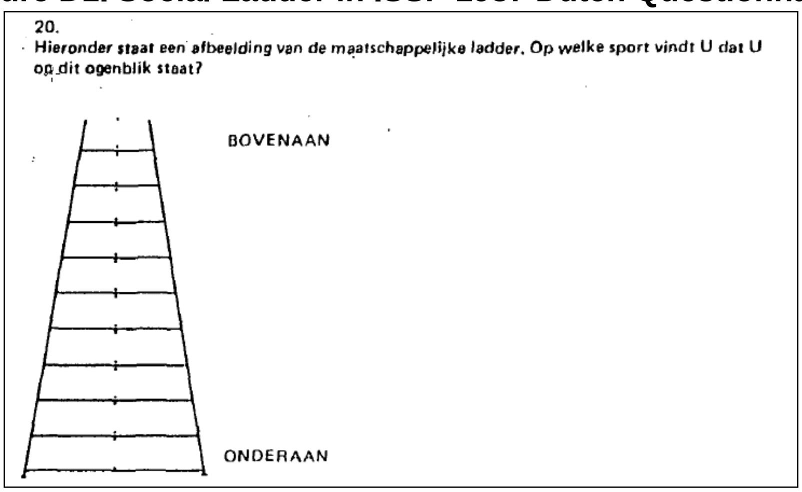


Figure D2. Social Ladder in ISSP 1987 Great Britain Questionnaire

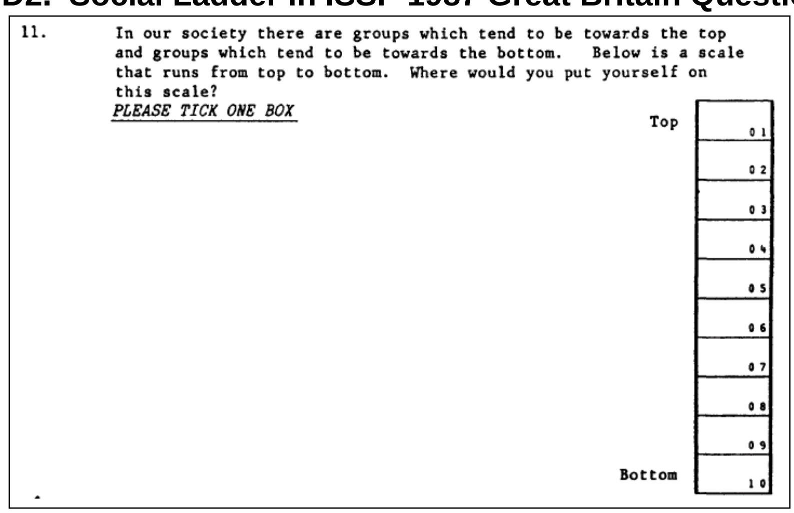


Figure D3. Social Ladder in ISSP 2004 Dutch Questionnaire

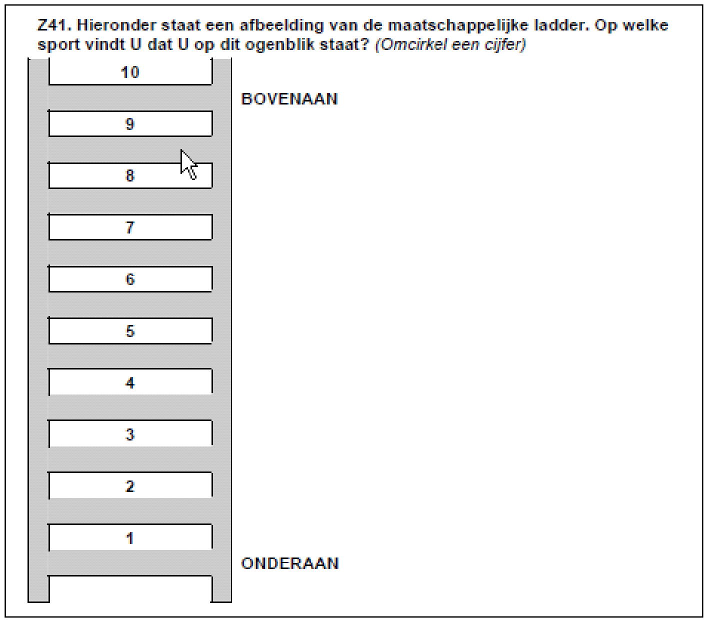

Z41. Hieronder staat een afbeelding van de maatschappelijke ladder. Op welke sport vindt U dat U op dit ogenblik staat? *(Omcirkel een cijfer)*

Figure D4. Ideal Shape Question in ISSP 2007 Austrian Survey

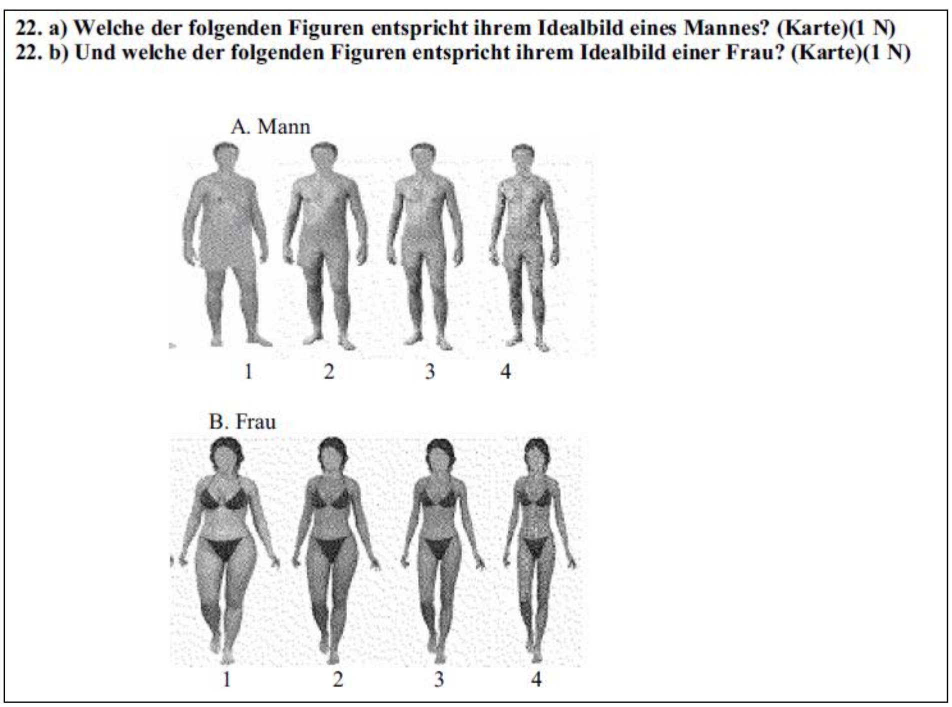
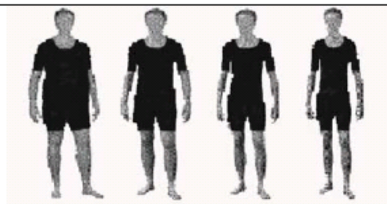

22. a) Welche der folgenden Figuren entspricht ihrem Idealbild eines Mannes? (Karte)(1 N)
22. b) Und welche der folgenden Figuren entspricht ihrem Idealbild einer Frau? (Karte)(1 N)

Figure D5. Ideal Shape Question in ISSP 2007 Phillipines Survey

71-72. Alin sa mga sumusunod na larawan ang pinakamalapit sa inyong palagay ng "ideal" na hugis ng lalaki at babae?
(PAKIBILUGAN ANG KAHON SA ILALIM NG LARAWAN NG PINILI NINYONG "IDEAL" NA HUGIS NG LALAKI AT BABAE.)

Which of the following pictures come closest to your conception of an ideal shape of a man and a woman?
(Please tick the box below the picture of your choice of the ideal man)

71. Man	72. Woman
	

Appendix E

Following are examples of text direction used by various countries in Asia, taken from the East Asia Barometer (EAB) survey in 2006 and the International Social Survey Programme (ISSP) in 2007. These suggest both that design guidelines need to reflect an understanding of how different Asian countries display text both vertically and horizontally, and that it would be desirable to pretest separately questions that differ across countries.

Figure E1. The 2006 EAB Singapore Questionnaire: Horizontal Response Option Column Headers Read from Top to Bottom

F. ACCESS TO PUBLIC SERVICE 公共服务

44-47. Based on your experience, how easy or difficult is it to obtain the following services? Or do you never try and get these services from government? (**Do not read**: Can't choose & Decline to answer)
根据您的经**验**, **请问**您**觉**得下**列**政府提供的公共服务, 容不容易**获**得? 或者, 您从未**获**得下**列**服务?

(SHOWCARD) 【**访员**出示卡片】	1 **Very Easy** 2 非常 容易	**Easy** 容易	**Difficult** 困**难**	**Very Difficult** 非常 困**难**	**Never Try** 从**来** **没**有	**Can't choose** 无法 **选择**	**Decline to answer** 不回答
44. An identity document (such as a birth certificate or passport) 申**办证**件服务 (例如身份**证**、护照等)	4	3	2	1	5	8	9
45. A place in public primary school for a child **为**小孩申**请**入学	4	3	2	1	5	8	9
46. Medical treatment at a nearby clinic 医**疗**服务 (在附近的医**疗诊**所看病)	4	3	2	1	5	8	9
47. Help from the police when you need it 要求警察帮助与服务	4	3	2	1	5	8	9

Figure E2. 2006 EAB Taiwan questionnaire: Vertical Response Option Column Headers

F. 公共服務

44-47. 根據您的經驗，請問您覺得下**列**政府提供的公共服務， 容不容易獲得? (或者， 您從未獲得下**列**服務)

	很容易	容易	困難	很困難	從來沒有	不適用	無法選擇	不回答	
	4.	3.	2.	1.	5.	0.	8.	9.	
(44) 申辦證件服務 (例如戶籍謄本、 護照等)	□	□	□	□	□	□	□	□	□68
(45) 為小孩申請入學	□	□	□	□	□	□	□	□	□69
(46) 醫療服務 (去附近的醫療診所看病)	□	□	□	□	□	□	□	□	□70
(47) 要求警察幫助與服務	□	□	□	□	□	□	□	□	□71

Figure E3. The 2006 EAB Mainland China Questionnaire: Vertical Response Option Column Headers, Read from Right to Left

I5. 基于您以往的经**验**, 您**觉**得**获**得下**列**服务有困**难吗?** 您曾经**尝试**从政府机构**获**得类似的服务**吗?**

	非常容易	容易	困难	很困难	从没试过	[不读] 无法选择	[不读] 不回答
I5a. 办理身份证、 出生证、 护照	4	3	2	1	5	8	9
I5b. 孩子在公立学校注**册**上学	4	3	2	1	5	8	9
I5c. 在附近医院看病或拿药	4	3	2	1	5	8	9
I5d. 在需要的**时**候**获**得警察的帮助	4	3	2	1	5	8	9

Figure E4. The 2007 ISSP Japan Questionnaire: Vertical Response Option Column Headers, Read from Left to Right

問３ あなたは、次にあげるＡ～Ｄの余暇活動を、どのくらい楽しんでいますか。最もあてはまるものに１つだけ○をつけてください。（**○はそれぞれ１つずつ**）

	1 まったく楽しんでいない	2 あまり楽しんでいない	3 まあ楽しんでいる	4 かなり楽しんでいる	5 非常に楽しんでいる	6 していない
Ａ．読書をる……………………………………………→	1……	2 ……	3 …….	4 ……	5 ……	6
Ｂ．友人とう……………………………………………→	1……	2 ……	3 …….	4 ……	5 ……	6
Ｃ．運動をする（スポーツをする、ジムに行く、散歩をする……………………………………→	1……	2 ……	3 …….	4 ……	5 ……	6
Ｄ．テレビ、ＤＶＤ、ビデオを見る………………→	1……	2 ……	3 …….	4 ……	5 ……	6

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Audit trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.
Behavior coding	Systematic coding of the interviewer-respondent interaction in order to identify problems and sometimes to estimate the frequency of behaviors that occur during the question-answer process.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: *Do you smoke?* *Yes* ___ *No* ___
Codebook	A document that provides question-level metadata that is matched to variables in a dataset. Metadata include the elements of a data dictionary, as well as basic study documentation, question text, universe statements (the characteristics of respondents who were asked the question), the number of respondents who answered the question, and response frequencies or statistics.
Coding	Translating nonnumeric data into numeric fields.

Cognitive interview	A pretesting method designed to uncover problems in survey items by having respondents think out loud while answering a question or retrospectively.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Computer assisted self interviewing (CASI)	A mode in which a computer displays the questions on a screen to the respondent and the respondent then enters his/her answers into the computer.
Consistency	Consistency is achieved when the same term or phrase is used throughout a translation to refer to an object or an entity referred to with one term or phrase in the source text. In many cases, consistency is most important with regard to technical terminology or to standard repeated components of a questionnaire. Reference to "showcard" in a source questionnaire should be consistently translated, for example. The translation of instructions which are repeated in the source text should also be repeated (and not varied) in the target text.
Context effects	The effect of question context, such as the order or layout of questions, on survey responses.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Data dictionary	A document linking the survey instrument (questionnaire) with the dataset, or more abstract question or variable-level metadata including question identifiers (variable names and labels); response category identifiers (value labels), and data types (e.g., F2.0, specifying that the response is a two-digit integer with zero decimal places.

Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Embedded experiments	Embedded experiments are included within the framework of an actual study.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Hard consistency check	A signal warning that there is an inconsistency between the current response and a previous response; the interviewer or respondent cannot continue until the inconsistency is resolved.
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Interface design	Aspects of computer-assisted survey design focused on the interviewer's or respondent's experience and interaction with the computer and instrument.
Interviewer effect	Measurement error, both systematic and variable, for which interviewers are responsible.
Item nonresponse, item missing data	The absence of information on individual data items for a sample element where other data items were successfully obtained.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.

Metadata Information that describes data. The term encompasses a broad spectrum of information about the survey, from study title to sample design, details such as interviewer briefing notes, contextual data and/or information such as legal regulations, customs, and economic indicators. Note that the term 'data' is used here in a technical definition. Typically metadata are descriptive information and data are the numerical values described.

Mode Method of data collection.

Nonresponse error Survey error (variance and bias) that is introduced when not all sample members participate in the survey (unit nonresponse) or not all survey items are answered (item nonreponse) by a sample element.

Open-ended question A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers.
Example: *What is your occupation?*
*Please write in the name or title of your occupation*___________

Paradata Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.

Precision A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.

Pretesting A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.

Progress indicator	An indicator that refers to aspects of reaching the goal (e.g., number of complete interviews).
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Question-by-question objectives	Text associated with some questions in interviewer-administered surveys that provides information on the objectives of the questions.
Response distribution	A description of the values and frequencies associated with a particular question.
Response options	The category, wording, and order of options given with the survey question.
Response styles	Consistent and stable tendencies in response behavior which are not explainable by question content or presentation. These are considered to be a source of biased reporting.
Sample element	A selected unit of the target population that may be eligible or ineligible.

Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Soft consistency check	A signal warning that there is an inconsistency between the current response and a previous response. The soft consistency check should provide guidance on resolving the inconsistency, but the interviewer or respondent may continue the survey without resolving it.
Source instrument	The original instrument from which other (target) instruments are translated or adapted as necessary.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample data.
Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Universe statement	A description of the subgroup of respondents to which the survey item applies (e.g., “Female, ≥ 45, Now Working”).

Usability testing Evaluation of a computer-assisted survey instrument to assess the effect of design on interviewer or respondent performance. Methods of evaluation include review by usability experts and observation of users working with the computer and survey instrument.

Variance A measure of how much a statistic varies around its mean over all conceptual trials.

XML (eXtensible Markup Language) XML (Extensible Markup Language) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere. XML documents are made up of storage units called entities, which contain either parsed or unparsed data. Parsed data is made up of characters, some of which form character data, and some of which form markup. Markup encodes a description of the document's storage layout and logical structure. XML provides a mechanism to impose constraints on the storage layout and logical structure.

References

[1] Aykin, N. (Ed.). (2005). *Usability and internationalization of information technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

[2] Aykin, N., & Milewski, A. E. (2005). Practical issues and guidelines for international information display. In Aykin, N. (Ed.), *Usability and internationalization of information technology* (pp. 21-50). Mahwah, NJ: Lawrence Erlbaum Associates.

[3] Christian, L. M., Dillman, D. A., & Smyth, J. D. (2005). Helping respondents get it right the first time: The influence of words, symbols, and graphics in Web surveys. *Public Opinion Quarterly 71*(1), 1-13.

[4] Clemmensen, T., & Goyal, S. (2005). *Cross-cultural usability testing—The relationship between evaluator and test user* (Working paper no. 06-2005). Frederiksberg Municipality, Denmark: Department of Informatics, Copenhagen Business School. Retrieved April 16, 2010, from http://openarchive.cbs.dk/bitstream/handle/10398/6474/06_2005.pdf?sequence=1

[5] Couper, M. P. (2002). *New technologies and survey data collection: Challenges and opportunities*. Paper presented at the International Conference on Improving Surveys, Copenhagen: Denmark.

[6] Couper, M. P. (2008). *Designing effective web surveys*. New York, NY: Cambridge University Press.

[7] Couper, M. P. (2001). *Web surveys: The questionnaire design challenge*. Paper presented at Proceedings of the 53rd session of the International Statistical Institute. Retrieved April 16, 2010, from http://isi.cbs.nl/iamamember/CD2/pdf/263.PDF

[8] Couper, M. P. (1999). The application of cognitive science to computer-assisted interview. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds) , *Cognition and survey research*. New York, NY: John Wiley & Sons.

[9] Couper, M. P., Beatty, P., Hansen, S. E., Lamias, M., & Marvin, T. (2000). *CAPI design recommendations* (Report submitted to the Bureau of Labor Statistics).

[10] Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly 65*(2), 230-253.

[11] de Leeuw, E., Hox, J., & Kef, S. (2003). Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods 15*(3), 233-251.

[12] Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York, NY: John Wiley & Sons.

[13] Dillman, D. A., Gertseva, A., & Mahon-Haft, T. (2005). Achieving usability in establishment surveys through the application of visual design principles. *Journal of Official Statistics 21*, 183-214.

[14] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[15] Hansen, S. E., & Couper, M. P. (2004). Usability testing as a means of evaluating computer assisted survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al (eds), *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley & Sons.

[16] Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review 21(3)*, 360-373.

[17] Hert, C.A. (2001). Studies of metadata creation and usage. *Proceedings of the federal committee on statistical methodology.* Washington, DC: U.S. Department of Commerce.

[18] Jagne, J., & Smith-Atakan, A.S.G. (2006). Cross-cultural interface design strategy. *Universal Access in the Information Society, 5*(3), 209-305.

[19] Kondratova, I., & Goldfarb, I. (2007). Color your website: Use of colors on the web. In N. Aykin (Ed.), *Usability and internationalization: Global and local user interfaces 4560* (pp. 123-132). Berlin, Germany: Springer.

[20] Murphy, E. D., Nichols, E.M., Anderson, A. E., Harley, M. D., & Pressley, K. D. (2001). *Building usability into electronic data-collection forms for economic censuses and surveys*. Paper presented at Federal Committee on Statistical Methodology Research Conference, Arlington, VA.

[21] Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web survey design. *Public Opinion Quarterly, 70*(4), 596-607.

[22] Russo, P., & Boor, S. (1993). How fluent is your interface?: Designing for international users. *Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems*. New York, NY: Association for Computing Machinery.

[23] Smith, T. W. (1995). Little things matter: A sampler of how differences in questionnaire format can affect survey responses. *Proceedings of the section on survey research methods* (pp. 1046-1051).

Alexandria, VA: American Statistical Association.

[24] Smith, T. W. (2005). The laws of studying societal change. *Survey Research, 36*(2), 1-5. Retrieved April 28, 2020, from http://www.srl.uic.edu/Publist/Newsletter/2005/05v36n2.pdf

[25] Sun, X., & Shi, Q. (2007). Language issues in corss cultural usability testing: A pilot sutdy in China. In N. Aykin (Ed.), *Usability and internationalization: Global and local user interfaces 4560* (pp. 274-284). Berlin, Germany: Springer.

[26] Survey Research Center. (2007). *SRC Blaise standards.* Ann Arbor, MI: The University of Michigan.

[27] Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883.

[28] Vatrapu, R., & Pérez-Quiñones, M. A. (2006). Culture and usability evaluation: The effects of culture in structured interviews. *Journal of Usability Studies, 1*(4), 156-170.

Internet Links

[29] European Social Survey (ESS). (2010). Retrieved May 03, 2010, from http://www.europeansocialsurvey.org

X. Interviewer Recruitment, Selection, and Training

Kirsten Alcser and Judi Clemens

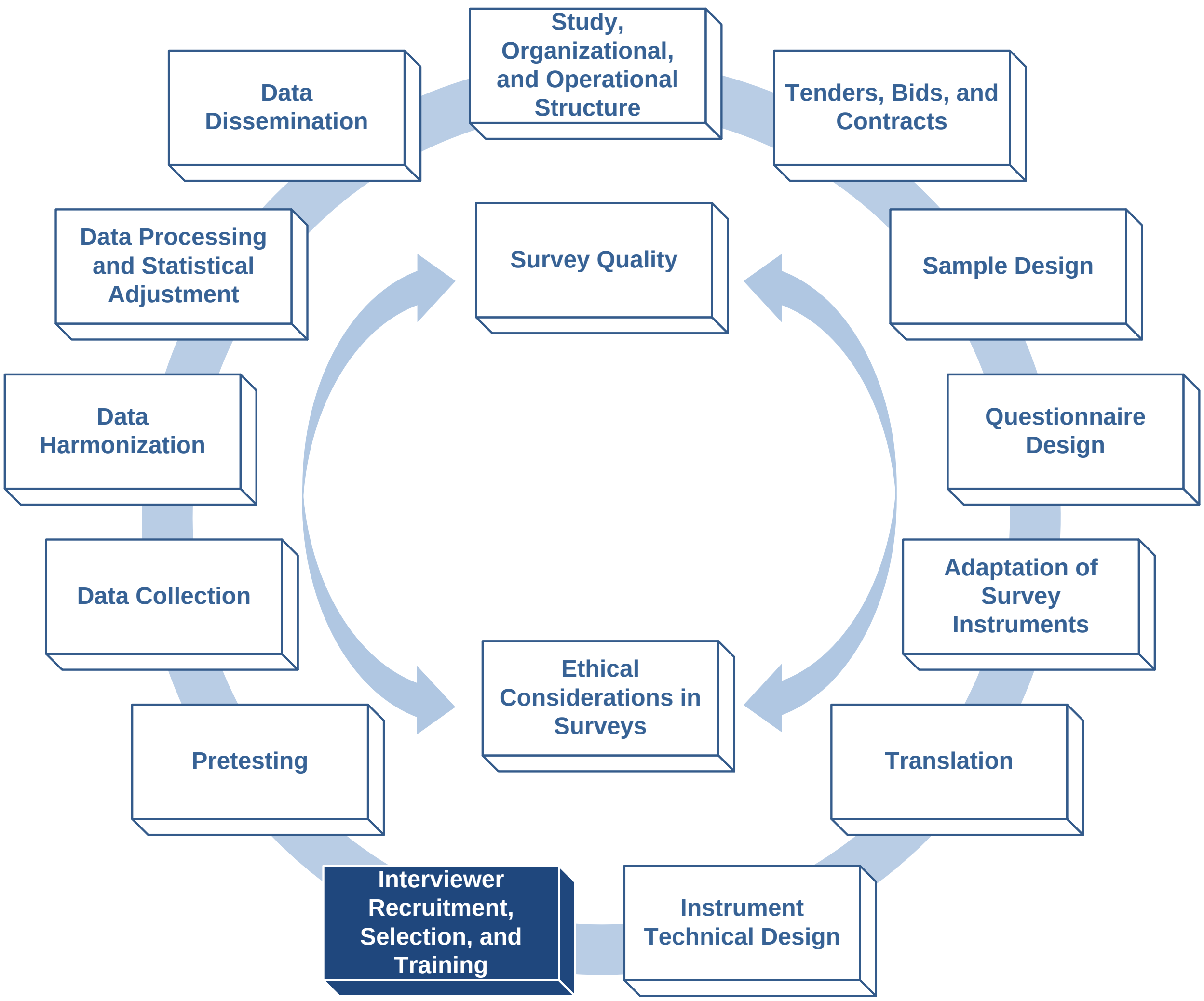
Introduction

Interviewers play a critical role in surveys, as they implement the survey design. They are often required to perform multiple tasks with a high level of accuracy. In a face-to-face household survey, the interviewer may be required to physically locate the household and to update the sample frame. In both telephone and face-to-face surveys, the interviewer has to contact the household, explain the purpose of the study, enumerate household members, select the respondent, motivate the respondent to participate, ask questions in the required manner, put the respondent at ease, and accurately record the respondent's answers as well as any other required information. Depending upon the survey topic and survey context, the interviewer may be required to perform additional tasks, such as bio-measure collection or oral translation.

Interviewers can influence responses through their personal attributes and their behaviors ("interviewer effects"). These guidelines present strategies to optimize interviewer efficiency and minimize the effect interviewer attributes have on the data through appropriate recruitment, selection, and case assignment; they also present strategies to minimize the effect that interviewer behaviors have on sampling error, nonresponse error, measurement error, and processing error through training.

Figure 1 shows interviewer recruitment, selection, and training within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To improve the overall quality of the survey data by minimizing interviewer effects while controlling costs by optimizing interviewer efficiency.

1. Determine the structure and composition of the interviewing staff.

Rationale

The structure and composition of the interviewing staff must be established during the design and planning phases of the project because these decisions will determine the number and type of interviewers required, the training protocol, sample assignment, and most efficient methods of supervision.

Procedural steps

- Consider such parameters as sample size and, for face-to-face studies, geographic distribution; the timing and duration of the data collection period; budget constraints; and the language(s) in which interviewing will occur [44].

- For face-to-face studies, decide whether interviewers will travel, either individually or in teams with a supervisor, or be locally assigned.
    - Factors favoring the use of traveling interviewers include:
        - Lower training costs compared to using local interviewers, as there are fewer interviewers to train and trainers do not have to travel to as many different locations.
        - Breach of confidentiality is less of an issue than with local interviewers because interviewers are unlikely to know the respondent personally.
        - Respondents may be more willing to participate in sensitive-topic surveys if the interviewers are strangers or "outsiders" [34].
    - Additional factors favoring the use of traveling teams rather than traveling individual interviewers include:
        - Traveling as a group may be safer than traveling individually.
        - Monitoring and supervision are easier than with individual traveling interviewers, since the supervisor is part of the group and is in close daily contact with the interviewers.
        - Interviewers have more opportunity to share experiences, learn from one another, and support one another than they would if traveling individually.
        - If multiple household members need to be surveyed, different interviewers can speak to them concurrently.
        - Similarly, if privacy is difficult to achieve, one interviewer can

speak to the respondent while another engages other household members.
    - It is easier to implement interpenetrated sample assignments than it would be with individual traveling interviewers [24].
  - Factors favoring the use of local interviewers include:
    - Employing a larger number of interviewers, each with a smaller workload, reduces the interviewer design effect [32] [40] (see Appendix A).
    - With a larger field staff, data collection can be completed within a shorter period of time, although the effect is not linear.
    - More call attempts can be made per case, since the interviewer remains in the area throughout the data collection period.
    - Local interviewer assignment reduces the need for interviewers to travel large distances, thereby reducing travel costs and time expended.
    - Local interviewers are familiar with the area and are more likely to share the language and customs of respondents; they may achieve higher response rates than would a stranger or "outsider."

- For telephone studies, decide whether interviewers will conduct the survey from a central telephone facility or from their homes (that is, decentralized telephone interviewing).
  - Factors favoring the use of centralized telephone interviewing include:
    - Training can be easily centralized.
    - Monitoring and supervision can be easier and less expensive, since the supervisor is in close daily contact with the interviewers and may, as a result, have access to more information of relevance.
    - It is easier to transfer sample elements among interviewers.
    - Cost controls are more efficient.
  - Factors favoring the use of decentralized telephone interviewing include:
    - A dedicated telephone facility is not required.
    - Interviewer working hours can be more flexible.
  - Some organizations already have a system in place which mixes centralized and decentralized telephone interviewing.
    - In these cases, retaining the combination of centralized and decentralized interviewing may minimize disruption and maintain flexibility.
    - Establishing a sample management system that pulls together information from the two into a single report can be a challenge.

- Estimate the Hours Per Interview (HPI). The HPI includes time spent traveling to all sample elements, attempting to contact them, documenting contact attempts, and working on project-related administrative duties, as well as conducting the interview with those respondents who agree to participate. The HPI, combined with the hours per week that each interviewer is expected to work on the project and the total number of weeks planned for data collection, helps determine the number of interviewers required (see Appendix B for an example).

- Utilizing the results of feasibility studies (see Data Collection), consider any special requirements of the study, such as:
  - How many languages are spoken and in what regions?
  - Are any specialized skills or knowledge required?
  - Would interviewer familiarity with the topic introduce bias or enhance an interviewer's ability to collect data?
  - Do cultural norms or the nature of the topic necessitate matching interviewers and respondents by gender, dialect, religion, race, ethnicity, caste, or age?
  - Is physical stamina a consideration (e.g., if interviewers will be required to walk, ride, or bicycle long distances) [38]?
  - Is the sample widely dispersed, making interviewer access to a car or reliable public transportation a consideration?
  - Is interviewer safety an issue?

Lessons learned

- Many organizations use a combination of interviewer assignment protocols. For example, they may hire local interviewers to make initial contact with sample households, select the respondent, and, if he or she is willing, administer the survey. Later in the data collection period, special traveling interviewers (for instance, experienced interviewers who have proven to be especially skillful at gaining cooperation or relating to particular types of respondents) can be brought in to persuade those selected individuals who have expressed a reluctance to participate. Alternatively, local interviewers might be hired in heavily populated areas while traveling interviewers are sent to more remote regions.

- If traveling teams of interviewers are used, the interviewer may not always be conversant in the respondent's language, and local interpreters may be needed to facilitate data collection. For example, the French Institut National d'Etudes Démographiques has collected data in several Bwa villages in Mali for over 15 years. Although French is the official language of Mali, most villagers speak only Boma, so interpreters were essential for collecting data. The interviewer was responsible for administering the questionnaire, while the interpreter's job was to act as a neutral intermediary between the interviewer and respondent, conveying the words and the concepts attached to them to the two speakers [46] (see Translation).

- Matching interviewer and respondent characteristics may improve cooperation but only appears to impact survey data quality if the topic of the survey is related to an identifiable and stable interviewer attribute.
  - Indonesian researchers felt that matching interviewers with respondents in terms of age, marital status, and child-rearing experience improved rapport and willingness to participate during in-depth interviews [42].
  - Several studies indicate that when the topic of the survey (e.g., racial attitudes or women's rights) is related to a fixed interviewer attribute (e.g., race or gender), the interviewer attribute can affect respondents' answers [15] [24] [27] [30] [47] [49].
  - If the topic of the survey is not related to a fixed interviewer attribute, matching the interviewer and respondent on the attribute does not appear to affect data quality. Axinn et al. [3] found that matching Nepalese interviewers and respondents by gender and ethnicity for a health survey did not decrease the number of technical errors and "don't know" responses or reduce incorrect information gathered during the interview.

- Attempting to match interviewer and respondent characteristics may strain the project's resources, particularly if this is not an established practice in the locale.

2. **Determine the pay structure for the data collection staff.**

Rationale

Since data collection staff quality has a major impact on the quality of the data collected, it is important to attract and retain the most qualified interviewers possible.

Procedural steps

- Interviewer pay structures vary greatly across cultures. Depending on local labor laws, set interviewer pay comparable to the pay for other jobs requiring similar skills, ideally adjusted for regional cost of living standards.

- Keep in mind local research traditions, the mode of the survey, and local labor laws. The two standard policies are to pay interviewers an hourly rate or to pay per completed interview [17] [44].

- Factors favoring payment per interview:
  - It is most feasible if each completed interview takes approximately the same amount of interviewer effort, as is more likely in a telephone survey [44].
  - It is easier to monitor and control interviewer costs than when paying by the hour [44] [52].
- Factors favoring an hourly rate:
  - It is most feasible if the effort to complete an interview varies widely, as is common in face-to-face surveys [33] [44].
  - Interviewers have less incentive to perform hurried, sloppy work or even to fabricate interviews than when paid per interview [44] [52].
  - Interviewers are less likely to focus on easy cases while neglecting those who are hard to reach or hard to persuade to participate than when paid by the completed interview [17] [44].
  - Interviewers may be more willing to spend time on other important tasks (e.g., completing a thorough screening interview and entering comprehensive, accurate contact attempt records) than when paid by the completed interview.

- When determining pay, consider the length and complexity of the interview, the expected difficulties of obtaining cooperation, and the amount of record-keeping demanded of the interviewer [17].

- Pay interviewers for time spent in training.

- Adjust the pay rate based on interviewer experience and any special skills they may possess and require (e.g., bilingual interviewers).

- Consider offering incentives for work above a certain target (e.g., response rate, contact rate, refusal conversion rate) as a way to keep interviewers motivated [17] [55].
  - Incentives can be extra pay, prizes, or special rewards.
  - Overreliance on interviewer incentives for completed interviews may give interviewers a reason to fabricate interviews [55].
  - Any bonus system must be perceived by the interviewers as being fair. For example, different sample assignments can vary considerably in the challenges they pose for interviewers [11].

Lessons learned

- Most survey organizations have a standard policy concerning pay arrangements (either paying per interview or paying by the hour) which they may be unwilling to change [11].

- If interviewers are paid by the interview instead of by the hour, they may rush the critical respondent-interviewer rapport-building process. It is especially important for face-to-face interviewers to spend the time necessary to develop this rapport so that respondents feel comfortable reporting honestly, as this leads to higher-quality responses. For example, when approaching a household, face-to-face interviewers need to conform to the culture's introductory customs, such as drinking tea or meeting elders [28].

- To discourage hurried, sloppy work when paying per interview, some organizations set a cap on the number of interviews that each interviewer is allowed to conduct in a day. Another strategy is to offer bonuses for high quality work. For example, set a basic pay per interview plus an additional 10% if the interviewer makes fewer than some predetermined number of errors. This requires the survey organization to have a monitoring system in place, which can distinguish between minor and more serious interviewer errors and can identify errors that cannot be attributed to the interviewer but rather to system factors, such as question wording and technology failures.

- In contrast to face-to-face interviewing, an experiment with telephone interviewers found that their productivity increased when they were paid per interview as opposed to being paid per hour [12].

3. Recruit and select an appropriate number of qualified interviewers.

Rationale

The quality of an interviewer-administered survey depends, to a large extent, on the quality of the interviewers and their supervisors. It is important, therefore, to recruit and select the best possible people for the job. In addition, selecting candidates who are well suited for the job may lead to lower interviewer turnover and reduced survey costs.

Procedural steps

- Recruit applicants.
    - Sometimes the interviewing component of the study can be subcontracted to an external survey organization with an existing pool of interviewers. At other times, the research team will have to implement outreach measures to find interviewers, such as asking local contacts for suggestions, placing flyers in strategic locations, or putting ads in local papers. In the latter case, recruitment and

training will take longer.
  - Keeping in mind any special considerations, as described in Guideline 1, target sources of potential interviewer candidates. Professionals, such as traveling nurses, can be a good source of interviewers for health studies; teachers, or others with substantive knowledge of the study topic, may also be good candidates.
  - Keep cultural norms and logistical factors in mind when recruiting interviewers. For example, it may not be acceptable in some cultures for young people (e.g., college students) to interview older persons or for women to interview men and vice versa. Similarly, persons with other jobs may not be available to work on the study at the times when respondents are most likely to be at home.

- Recruit more than the number of interviewers needed for data collection to allow for attrition and the dismissal of candidates who prove to be unsuitable.

- As appropriate, prepare an application form to use in prescreening interviewer candidates before they are invited to an in-person or telephone job interview.

- Interview applicants in the mode of the study. That is, hold telephone screening interviews for a telephone survey and face-to-face screening interviews for a face-to-face study.

- Evaluate each candidate.
  - If appropriate for the culture, conduct a criminal background check, particularly if the interviewers will handle sensitive information or come into contact with vulnerable populations (e.g., the young, the old, the infirm).
  - Criteria for employment commonly include interviewing skills, language skills, computer or technical skills, organizational skills, education, availability, location, the ability to meet production (i.e., data collection) goals, and the capability to handle potentially emotional or stressful interactions with respondents [44].
  - When possible, select interviewers who have previously worked on similar studies and have good recommendations based on their performance. Experienced interviewers require less training and are likely to achieve higher response rates [11] [19].
  - Evaluate the accuracy and clarity with which each potential candidate can read and process the survey questions in the language(s) of the interview and make sure that he or she is comfortable reading out loud. Ideally, language proficiency should be formally assessed by an outside expert or language assessment firm and should include evaluation of [43]:

        - Conversational skills (e.g., comprehension level, comprehension speed, speech level, speech speed, and accent)
        - Writing skills (e.g., grammar, spelling, and the ability to enter responses)
        - Reading skills (e.g., reading aloud)
    - Ideally, test applicants' computer skills for studies using a computerized questionnaire and sample management system.
    - Select interviewers who are punctual and have good organizational skills (e.g., are able to handle forms and keep track of paperwork).
    - Select interviewers who have completed the full period of required schooling within their country.
    - For face-to-face studies, assess applicants' ability to read maps.

- Give the candidates a realistic preview of the job including the survey topic and the type of questions that will be asked; describe any non-traditional interviewing tasks (e.g., taking physical measures) in the recruitment description and the screening interview.

- Clearly present the candidates with study expectations for workload (weekly, monthly, including evening work and possibly weekend work).

- Obtain the candidates' written commitment to work at the expected level of effort for the duration of the data collection period.

- Base selection on an objective evaluation of the candidate's abilities rather than his or her relationship to survey staff or favoritism [38] [54] [57].

Lessons learned

- Vaessen et al. [54] suggest that study managers recruit at least 10-15 percent more than the number of interviewers ultimately needed for fieldwork to allow for attrition and the dismissal of candidates who prove to be unsuitable.

- A variety of selection criteria have been used successfully by established cross-cultural studies. In the Afrobarometer Survey, interviewers (preferably women) usually hold first degrees in social sciences and have some university education, strong facility in local language, and the ability to relate to respondents in a respectful manner (selection is on a competitive basis which may include reading, speaking, and comprehension of national and local languages, and competence at following detailed instructions) [57]; the Asian Barometer recruits interviewers from university graduates, senior social science undergraduates, and professional survey interviewers [58]; the European Social Survey highly recommends using experienced interviewers [59]; the Living Standard Measurement Study Survey requires that interviewers have completed secondary education and recommends fluency in two or more languages [36]; the Survey of Health, Aging and Retirement in Europe selects survey agencies for all participating countries and requires interviewers to have extensive face-to-face experience [60]; and in the World Mental Health Survey, some participating countries use field staff from established survey organizations, while others recruit new interviewers from the general population or among college students (interviewer criteria vary among participating countries and may include interviewing experience, language skills, technology skills, education, and capability to handle potential sensitive situations with respondents) [31].

- Liamputtong, a professor in the School of Public Health at La Trobe University, believes that bicultural researchers who are familiar with both the local and mainstream cultures of communities in the study are ideal [35].

- As noted in Guideline 1, it is not always possible to recruit interviewers who are fluent in the language(s) preferred or needed by respondents. In this case, other arrangements must be made. Options may include working with interpreters, data collection by proxy, using a bridge language if available, or using self-completion modes if literacy levels permit.

- If the topic is sensitive (e.g., domestic violence), empathy and strong interpersonal skills may be more important than high levels of education or previous interviewing experience [29]. This holds true for both interviewers and any interpreters being used.

- If the project's interviewing protocol differs significantly from previous studies, experienced interviewers may find it difficult to change their habits ("veteran effects"). In this case, it may be preferable to recruit and train new interviewers. Similarly, interviewers who have worked for an organization with low quality standards may have to unlearn some behaviors and adapt to new standards.

4. **Provide general basic interviewer training.**

Rationale

Newly hired interviewers and supervisors require basic training in techniques for successful interviewing before they receive specific training on the study on which they will be working. Research indicates that interviewer training helps improve the quality of survey data by: (1) reducing item nonresponse [8], (2) increasing the amount and accuracy of information obtained [8], and (3) increasing survey participation by teaching interviewers how to identify and respond to respondents' concerns [39].

Procedural steps

- Allow sufficient time to adequately cover general interviewing techniques material.

- Select appropriate trainers. These may include research staff, project managers, project management assistants, supervisors who directly oversee data collection staff, and experienced interviewers.

- Provide the following general information:
  - An overview of the organization.
  - The roles of the interviewer and the supervisor in the research process.
  - The format of the survey interview.
  - An overview of different interview modes (face-to-face, telephone, computer-assisted, observing behaviors and events, and delivering self-administered survey materials such as diaries) and the tasks each poses for the interviewer.
  - An overview of different sampling designs and the tasks each poses for the interviewer.
  - Interviewer evaluation procedures and criteria.

- Include the following prescribed procedures [20]:
  - Standardized question-asking. Train interviewers to read each question exactly as written and to read the questions slowly. They should ask all questions exactly in the order in which they are presented in the questionnaire [16] [24] (see Guideline 5 for exceptions).
  - Questionnaire format and conventions. Teach interviewers how to enter the answers to both open- and closed-ended questions. Train them to follow interviewing conventions such as emphasizing words in the questionnaire which appear in bold or are underlined, recognizing and not reading aloud interviewer instructions, reading or not reading optional words as appropriate, and selecting correct fill choices (e.g., he/she, has/have, etc.).
  - Clarification. If the study staff has not prepared a stock definition, train interviewers to repeat all or a specified part of the question verbatim when respondents ask for clarification. Interviewers

should not make up their own definitions to any word, phrase, or question in the questionnaire [11]. Train interviewers to notify their supervisors about any questions which are confusing to respondents and require further clarification.

- Probing. If a respondent's answer is inadequate and it is legally and culturally permissible to probe, train interviewers to employ unbiased techniques to encourage answers that are more complete, appropriate, and thoughtful [11] [24].
  - Such strategies of probing for more information may include a pause to encourage the person to fill the silence or a direct request for further information.
  - Verbal probes should be chosen from a stock list of phrases such as "Could you explain what you mean by that?" or "Can you tell me anything else about ___________ ?"
  - Stock phrases must be neutral; that is, they must avoid "sending a message" about what is a good or a bad response.
- Feedback. Train interviewers to provide their respondents with culturally appropriate feedback when they are doing well in order to encourage them to listen carefully and to give thoughtful answers [11].
  - This feedback may be in the form of a nonverbal smile or nod or a short encouraging phrase.
  - Verbal feedback should be selected from a prepared list of stock phrases such as "That's useful information" or "Thank you, that's helpful" to ensure that the feedback is not evaluative of the content of the answer. For example, in English the word "okay" is discouraged for use in feedback because it could be construed as agreement with or approval of the respondent's answer.
  - As a general rule, give nonverbal or short feedback to short answers and longer feedback phrases to longer answers.
- Recording answers. To reduce measurement error, train interviewers to record answers exactly as given.
  - If the question offers fixed alternatives, teach interviewers to get respondents to choose one of the fixed alternatives; interviewers should not infer which alternative is closest to what the respondent actually says [24].
  - If the question requires a narrative response, teach interviewers to record the answer in as near verbatim form as possible [24].
- Confidentiality. Train interviewers to keep confidential all identifying respondent contact information as well as respondents' answers to survey questions.
- Computer Assisted Personal Interviewing (CAPI) conventions (see Instrument Technical Design).
- Completing contact attempt records. Teach interviewers to record when each contact was attempted, pertinent respondent comments

(e.g., the best time to reach him or her or reasons for reluctance to participate), and the result of each contact attempt, using disposition codes (further information on contact attempt records and disposition codes can be found in Tenders, Bids, and Contracts and Data Processing and Statistical Adjustment; examples of contact attempt records can be found in Data Collection).

- Recording time and meeting production goals. Teach interviewers how to record the time they spend on each defined aspect of their work for the study, both for their remuneration and to allow supervisors to monitor their progress and efficiency during data collection.

- If legally and culturally permissible, teach interviewers noncoercive persuasion techniques and practice counter replies to common statements of reluctance.
  - Discuss optimal times and modes for contacting target persons.
  - Train interviewers to tailor their initial interactions with respondents by developing the following skills [25] [39]:
    - Learning the classes of concerns ("themes") that respondents might have.
    - Classifying the respondent's wording into the appropriate theme.
    - Addressing the concern, using their own words.
  - Employ hands-on practice exercises so that the trainees become proficient in quickly identifying respondent concerns and quickly responding to them.

- For best overall results, employ a training format that combines lecture with visuals and small-group practice sessions.
  - Mixing the format keeps the trainees engaged and acknowledges that different people learn in different ways [21].
  - Through practice, trainees move from procedural knowledge (knowledge of how to perform a task) to skill acquisition (the ability to perform the task almost automatically) [39].
  - Although the class can be large for lecture sessions, trainees should break up into smaller groups for hands-on practice.

- Be sensitive to the local culture.
  - Educate trainers in cultural sensitivity.
  - Take religious holidays into consideration when scheduling training sessions.
  - Make every effort to accommodate dietary restrictions when planning meals or snacks for the training.
  - Be aware that conventions regarding breaks vary among cultures.

- At the end of basic interviewer training, evaluate the knowledge of the interviewer candidates. This can be done by written test, conducting a scripted certification interview with a supervisor, audio taping, or observing the interviewer conduct an actual interview.

Lessons learned

- If the interviewer candidates have access to the necessary equipment, some basic interview training material can be presented in the form of audio- or video-recordings for home study [10] [18]. Other training options include telephone and video conferencing and self study using paper materials.

- A forthcoming article [56] indicates that interviewer-related variance on survey items may be due to nonresponse error variance rather than measurement difficulties. That is, different interviewers may successfully contact and recruit respondents with different characteristics (e.g., age, race), even though their sample pools start out the same.

- Interviewer training and the interviewer manual need to be adjusted to be culturally sensitive to the population under study:
  - Textbook instructions on handling reluctance to participate and provide accurate information rely to a large extent on Western experiences. When possible such procedures should be modified so that they include culturally acceptable and suitable tactics. Researchers conducting a women's health study on the Apsáalooke native American reservation in southeastern Montana, U.S.A., felt that standard Western tactics for handling reluctance would be offensive in that culture. They therefore did not attempt to persuade reluctant respondents to participate. In addition, interviewers were encouraged to display a compassionate attitude and interest in the women, rather than the standard recommended neutral voice tone and lack of responsiveness to respondent answers, to minimize eye contact, and to accept offers of food and drink – all to be more consonant with the Apsáalooke culture [13].
  - In majority countries, a Western trainer may be respected but resented. Researchers in Puerto Rico found allowing interviewer trainees to provide input about the local culture and supplementing trainer criticism with peer criticism helpful [50].
  - The World Mental Health study added country-specific topics to their general interviewer training sessions. In New Zealand, they included cultural empathy to Maori and Pacific Islander households; in Colombia, they provided special training on interacting with governmental authorities and armed guerrilla and paramilitary

groups [44].

5. **Provide study specific training for all interviewers and supervisors.**

Rationale

Interviewers and supervisors need to be very familiar with the study's protocols and questionnaire in order to carry out their tasks. Depending upon the survey, they may need to learn the instrument's branching structure, the study's requirement for field coding, or the use of a respondent booklet, show cards, or other visual materials. There may be special instructions for implementing all or part of the survey that deviate from the standardized interviewing covered in general interviewer training. Interviewers should also be knowledgeable about the project objectives so that their actions help, not hinder, the overall goals. Both newly hired and experienced interviewers require training specific to the study at hand.

Procedural steps

- Allow sufficient time for study-specific training, depending upon the complexity of the study (see Appendix C for a sample training agenda).

- When possible, have the same team from the coordinating center train all interviewers to ensure standardization of study-specific protocols [53] (see Study, Organizational, and Operational Structure for more on the coordinating center). The team may provide regional trainings, traveling to where interviewers are located.

- Select appropriate trainers. These may include research staff, project managers and people on their staffs, supervisors who directly oversee data collection staff, experienced interviewers, and consultant(s) hired to assist with interviewer training.

- Include a large amount of practice and role playing using the questionnaire [53].
  - Consider having the interviewers complete a self-interview to become familiar with the survey instrument.
  - Hands-on training may include round-robin practice sessions (i.e., scripted practice sessions where interviewers take turns administering survey questions to the trainer in a group setting), mock one-on-one interviews (i.e., sessions where interviewers interview each other), listening and discussing taped interviews, and live practice with potential respondents.
  - For role playing to be effective, prepare different scripts in advance

so that the different branching structures of the interview, the nature of explanations that are permitted, and anticipated problems can be illustrated.
    - Consider making a video to illustrate the correct administration of physical measures, if applicable. This ensures that the material is consistently taught, especially if training is conducted at multiple times or in various locations.

- Provide interviewers with an Interviewer Project Manual/Study Guide that has been prepared by the coordinating center, with input from local collaborators. The manual is an important part of training and will serve as reference material while the survey is underway [22].
    - Complete and review the manual before training begins [22].
    - When appropriate, translate the manual into the languages used in the geographical areas encompassed by the study.
    - Include the following content in both the training agenda and the project manual:
        - General information about the project (e.g., the study's background and goals, funding sources, and principal investigators).
        - How to introduce the survey to respondents.
        - Eligibility and respondent selection procedures, if applicable. Sampling and coverage errors can occur if interviewers fail to correctly locate sample households, determine eligibility, or implement the respondent selection procedure [37].
        - Review of the survey instrument, highlighting the content of the various sections and the types of questions being asked.
        - Data entry procedures (paper and computer-assisted). Measurement error can occur if interviewers do not record responses in the appropriate manner.
        - Computer hardware and software usage, if appropriate (e.g., use of the laptop computer, email, and any other software packages).
        - Use of the sample management system.
        - Review of interview procedures and materials (e.g., informed consent materials and respondent incentive payments).
        - Review of study-specific probing conventions (e.g., when to probe a "don't know" response and an open-ended response).
        - Techniques for handling reluctance that are specific to the study (e.g., recommended responses to frequently asked questions) and are approved in advance by an ethics review committee (see Ethical Considerations in Surveys). Nonresponse bias can occur if interviewers are unable to persuade reluctant persons to participate in the survey.
        - Nonstandardized interviewing, if appropriate for the study (e.g., event history calendars, time diaries, or conversational

interviewing) [5] [6] [14] [24] [51]. (See Data Collection for a discussion about combining qualitative and quantitative data collection methods.)
- Any observational data which interviewers will be required to enter (e.g., observations of the respondent or the neighborhood).
- Any specialized training for the study (e.g., procedures for taking physical measurements, instruction on interviewing minors or interviewing on sensitive topics, proxy interview protocol, interviewing in unsafe neighborhoods, and protocol for handling respondent or interviewer distress).
- Procedures to be used for unusual cases, including general principles to be applied in dealing with unforeseen problems (e.g., how to report abuse of children or others that is observed while conducting an interview in the respondent's home).
- Production goals and maintaining productivity.
- Proper handling of equipment and survey materials.
- The structure of the survey team and the role of all members of the team.
- Procedures for editing and transmitting data. Processing error can occur if interviewers do not correctly edit and transmit the completed questionnaire (see Data Processing and Statistical Adjustment for other potential sources of processing error).
- Any other required administrative tasks.

- The Project Manual/Study Guide must be especially clear and self-contained if it is impossible to train interviewers in person (e.g., if interviewers must be trained via conference call or video).

- Collect and analyze written evaluative feedback (i.e., provide the opportunity for trainees to give written feedback on trainer performance, the sufficiency of time allocated to different topics, and the adequacy of practice exercises).

- Certify the interviewers. (See Appendix D for a sample interviewer certification form.) Certification for study-specific tasks should include:
  - A complete role-play interview with a supervisor.
  - Certification by an appropriate trainer of any physical measurements that are included in the study (see Appendix E for a sample certification checklist for taking physical measurements).
  - Language certification, as appropriate (see Translation).

- Supplement the initial training with periodic in-person seminars, telephone conference calls, and periodic bulletins or newsletters [44].

- If data collection will extend for a long period of time, hold a brief refresher training course towards the middle of the data collection period [40].
  - This refresher training session is an opportunity to review various aspects of data collection, focusing on difficult procedures or on protocols that are not being adhered to sufficiently by interviewers.
  - The session can also be used to provide feedback on what has been achieved to date.
  - Require even experienced interviewers to attend refresher training sessions, including sessions on standardized techniques.

Lessons learned

- Most of the time it is not feasible for the same team to train all interviewers, particularly in very large cross-cultural studies. If this is the case, other steps must be taken to ensure the standardization of study-specific protocols:
  - One approach is the “train-the-trainer” model.
    - Training is generally done in one common language
    - Each country or cultural group sends one or more individuals, who can understand and work in the language of the trainers, to the central training.
    - These representatives return to their own country or cultural group, adapt and translate the training materials as needed, and train the interviewers.
    - This model allows for tailoring at the country or cultural group level.
    - The Survey of Health, Ageing, and Retirement in Europe (SHARE) train-the-trainer (TTT) program is one example of this approach [1] [2] [9] [60]. The University of Michigan’s Survey Research Center, under contract to SHARE, created the TTT program. Each participating country sent a Country Team Leader, a member of his or her staff, and 2-3 trainers to the TTT sessions. Once the trainers had completed the TTT program, they used the training materials provided, translated if necessary, to conduct country-level interviewer training (see Appendix C for the SHARE Model Training Agenda).

    - The World Mental Health Survey gives two train-the-trainer sessions for interviewer supervisors, lasting, on average, six days. Interviewer supervisors in turn, train the interviewers in general interviewing techniques (on average 20 hours) and CIDI specific training (on average 30 hours). Before progressing to CIDI specific training, interviewers must demonstrate competence, in the form of role playing, tests, and/or supervised respondent interaction, and in general interview techniques. All interviewers must be tested and certified before they are authorized for production work [31].
  - Another approach is the training center model [45].
    - A centralized training course is held, but language “regions” are represented rather than countries.
    - This model is effective when it is not possible for every country to send trainers who are functional in the central trainer’s language.
    - The training center model was used in the World Health Organization’s Composite International Diagnostic Interview training sessions. For example, trainers from Lebanon were trained in the United States and subsequently trained the trainers in Lebanon, Oman, Jordan, Palestine, Saudi Arabia, and Iraq.
  - Organizing training in steps (first training the trainers and then having them train the interviewers) increases the overall time needed for training and should be factored into the project timeline.
  - All step-wise training results in a certain loss or distortion of information as it is passed along. Trainers should be aware of this and take precautions, such as providing approved standardized training materials.

- If interviewers are being hired for one study only, basic interviewer training techniques can be incorporated into study-specific training.

- The amount of time devoted to training varies among large established cross-cultural surveys. Glewwe [22] recommends up to a month of intense interviewer training (general and study specific) for inexperienced interviewers in a face-to-face survey. Field team members for the Asian Barometer received intensive, week long training sessions on the questionnaire, sampling methods, and the cultural and ethical context of the interview [58]. The Living Standard Measurement Study Survey recommends that training take place over a four week period and include introduction to the LSMS survey, general survey procedures, the questionnaire, sampling procedures, and data entry program error reports, with at least two observed training interviews [36]. The Survey of Health, Ageing, and Retirement in Europe (SHARE) requires 16-18 hours of training spread over 2-3 days in addition to the basic interviewer techniques training for new interviewers [1] [2]. Similarly, the World Health Survey (WHS) recommends three full days of study-specific training [53].

- Round 4 of the Afrobarometer Survey held a six day training workshop for all persons involved with the project, including interviewers and field supervisors. The Afrobarometer protocol requires holding a single national training workshop at one central location. Interviewers must complete at least six practice interviews before they leave for the field: at least one mock interview in the national language, at least one mock interview in each of the local languages they will use in the field, and at least four training interviews in a field situation [57].

- In addition to general interview training, all interviewers for round 5 of the European Social Survey were briefed by the National Coordinator or a research team member regarding respondent selection procedures, registration of the calling process, response rate enhancement, coding of observation data, documentation, and questionnaire content. Practice interviews were suggested [59].

- If the topic is extremely sensitive, additional specialized training may improve response rates and data quality. The WHO Multi-Country Study on Women's Health and Domestic Violence, fielded in multiple culturally diverse countries, found that previously inexperienced interviewers who had received specialized training obtained a significantly higher response rate and significantly higher disclosure rate of incidences of domestic violence than did experienced interviewers who had not received the additional training [29].

- Training interviewers in adaptive behavior, such as tailoring responses to respondent concerns or nonstandardized conversational interviewing, can be time-consuming and could increase training costs [39].

- Field interviewers often work some distance away from their trainers and supervisors. Before sending the interviewers to their assigned areas, some organizations have found it useful to have them conduct a few interviews close to the training locale. Afterward, they meet with the trainer, discuss their experiences, and check their questionnaires. Any problems or misunderstandings can be identified and rectified more easily than if they had occurred in a more remote area.

- During pretesting for the Tamang Family Research Project, investigators trained interviewers in a Nepalese village that was not in the sample. The investigators and interviewers lived together during this period and throughout data collection. This allowed for the continuous assessment of interviewers who were let go if they were not completing quality work [4].

6. **Institute and follow appropriate quality control measures.**

Rationale

- Quality control (QC) is a procedure or set of procedures intended to ensure that a product or service adheres to a defined set of quality criteria or meets the requirements of the study (see Survey Quality). The implementation of quality control measures enhances the accuracy, reliability, and validity of the survey data and maximizes comparability of these data across cultures. To implement an effective QC program in a cross-cultural survey context, the coordinating center must first decide which specific standards must be met. Then real-world data must be collected and the results reported back to the coordinating center. After this, corrective action must be decided upon and taken as quickly as possible. Finally, the QC process must be ongoing to ensure that remedial efforts, if required, have produced satisfactory results.

Procedural steps

- Track the cost and success rates of different recruitment avenues to determine which are the most fruitful and cost effective; use this information to guide the future allocation of resources.

- Considering the factors enumerated in Guideline 3, establish a checklist of minimum interviewer candidate requirements (e.g., interviewing skills, reading/writing fluency, language skills, educational level, and computer skills).
  - Require recruiters to complete the checklist as they screen each interviewer candidate. If specific assessment tests are used (e.g., to evaluate language skills), record each candidate’s performance on the test.
  - Accept only those candidates who meet the predetermined minimum requirements.
  - To ensure accountability, require the recruiter to sign or initial checklists and assessment tests.

- Survey interviewer candidates to determine what improvements could be made to the recruitment process; use this information to modify the procedure, if possible (for example, ask how the candidate heard about the position).

- Take attendance at general interviewing techniques and study-specific training sessions.
  - Dismiss candidates who fail to attend a predetermined minimum number of training sessions, or make arrangements to train them individually on the missed material.
  - Keep a signed written record of the training completed by each candidate.

- At the end of basic interviewer training, evaluate the knowledge of the interviewer candidates, as described in Guideline 4.
  - Require all trainers to use the same evaluation criteria.
  - Dismiss or retrain those candidates who fail to attain predetermined minimum standards.
  - Keep a signed written record of each candidate's performance on the evaluation measures.

- At the end of study-specific training, certify the interviewer candidates, as described in Guideline 5.
  - Require all trainers to use the same evaluation criteria for certification.
  - Dismiss or retrain those candidates who fail to attain predetermined minimum standards.
  - Keep a signed written record of each candidate's performance on the certification tests.

- Debrief interviewer trainees to determine how training could be improved; use this information to modify the training protocol, if possible.

Lessons learned

- Including quality control protocols as part of the overall survey design, and implementing them from the start, permits the survey organization and the coordinating center to monitor performance and to take immediate corrective action when required. For example, if many interviewer candidates fail to pass the study-specific certification test, additional training could be provided. Afterward, the candidates would be tested again. Those passing the certification test could then be sent out into the field.

7. Document interviewer recruitment and training.

Rationale

Comprehensive documentation helps analysts correctly interpret the data and assess data quality; it also serves as a resource for later studies.

Procedural steps

- Document the recruitment effort for enrolling data collection staff on the project, including:
  - Any special criteria used in reviewing data collection staff employment applications (e.g., language proficiency and special knowledge and skills, such as taking physical/biological measurements).
  - The way in which language fluency was assessed, as appropriate for the study.
  - Recruitment scripts and sources used to recruit data collection staff, as well as an evaluation of the success of the recruitment strategies.
  - Interviewer characteristics (e.g., gender, age, race, education, length of tenure as interviewer).
  - Characteristics of the multilingual interviewing staff in terms of the percent certified to interview by language.
  - The minimum number of hours required, if applicable, and the average number of hours worked by an interviewer during the data collection period.
  - Interviewer pay structure (e.g., hourly or per completed interview), the pay range, and any bonus program (e.g., amount and when or under what circumstances these bonuses were offered).

- Document the general and study-specific training, including:
  - Number of training sessions conducted.
  - Number of training hours, dates, and locations.
  - Number of trainers and trainees.
  - Background of the trainers, including expertise in training and in any substantive areas as applicable to the survey.
  - Copy of the training agenda(s) (i.e., list of topics covered).
  - All written materials that were used (e.g., the interviewer manual/study guide, trainer/facilitator guide and supplemental training materials).
  - Certification procedures (e.g., scripted certification interview with a supervisor or other staff, written or online test on general interviewing procedures, live practice interviewing with potential respondents).

- Document any issues encountered (e.g., if the recruitment plan failed to produce a sufficient number of qualified interviewers or interviewer attrition was unexpectedly high, necessitating a second round of recruitment and training; the training agenda did not provide adequate time for hands-on practice; or the ratio of trainers to trainees was inadequate) and suggestions for future studies.

- Document all direct measurements of data quality, all indicators of data quality obtained via quality control (QC), and any decisions made to change the protocol in order to maintain high levels of quality (see Survey Quality).

Lessons learned

- Documenting the recruitment effort, including method(s) of recruiting, number of candidates recruited, and number of candidates screened, as well as post-study documentation of interviewer retention, is also useful for other future projects. This information can guide future recruitment strategies and help estimate the number of recruits needed to provide a sufficient number of interviewers for data collection in similar studies.

- Documentation of general and study-specific training can pinpoint areas needing improvement in future training efforts.

Appendix A

Interviewer Design Effect [24] [32] [40] [41]

Research indicates that the interviewer design effect may be even larger than the design effect attributable to geographic clustering [48]. This is especially true in some international studies where cultural and other factors contribute to large interviewer variances. Interviewer variance occurs when response errors of persons interviewed by the same interviewer are correlated; therefore interviewer variance is part of the correlated variance component of the total variance (other correlated variances stem from coders, editors, supervisors and crew leaders).

The intra class coefficient, ρ_{int}, is a measure of the ratio of interviewer variance to the total variance and is defined as:

$$\rho_{int} = \frac{(\text{between-interviewer variance})}{(\text{between-interviewer variance}) + (\text{within-interviewer variance})}$$

The value of ρ_{int} is theoretically always between 0 and 1 although calculated estimates of ρ_{int} may sometimes be negative. In this case, they are usually treated as zeros. When ρ_{int} for a particular variable is 0 or is negative, we interpret this to mean that the interviewers have no effect on the variance of responses to that variable; the larger the value of ρ_{int}, the larger the effect of interviewers on the variance of the particular variable.

The interviewer design effect ($deff_{int}$) is a measure of the effect of interviewers carrying out multiple interviews, compared to what you would get if there was a different interviewer for each respondent, all else being equal (if the addition of more interviewers increases costs such that supervision or training must be reduced to compensate, interviewer variance may actually increase).

$$deff_{int} = 1 + \rho_{int}\,(m-1)$$

where m is the average number of interviews per interviewer.

Thus, even a small interviewer variance (ρ_{int},) can have a significant effect on the variance of a survey estimate if m is large. The interviewer variance contribution is usually not included in textbook variance estimation formulas. Interviewer variance leads to a loss of sample information when the effective sample size n_{eff}, defined as $n/deff_{int}$, is smaller than the actual sample size n.

Standardized interviewing aims to reduce interviewer variance.

For specification of a mathematical model of response errors when interviewers are used, see [26]; for further discussion of interviewer variance see [7] and [23].

Appendix B

Estimating the number of interviewers needed for a study

The following example shows how to calculate the number of interviewers required for a hypothetical study. The example makes the following assumptions:

1. Interviewers and respondents do not need to be matched on any attributes.
2. The average number of hours worked per week is the same for all interviewers.
3. The expected number of completed interviews is 500.
4. The estimated Hours Per Interview (HPI) is 5.
5. The projected data collection period is 5 weeks.
6. Each interviewer is expected to work 15 hours per week (based on the optimal hours of work during the times the respondents are expected to be at home).

Make the following calculations:

1. Total hours to complete the study = (500 interviews * 5 HPI) = 2500 hours.
2. Average interviewer hours per week = (2500 total hours/5 weeks) = 500 hours per week.
3. Number of interviewers needed = (500 hours per week/15 hours per interviewer per week) = 33 interviewers.

(To determine the optimum number of interviewers based on interviewer variance and cost, see [26]).

Appendix C

Example of a Training Agenda [2]

The Survey of Health, Ageing and Retirement in Europe (SHARE) utilizes a model agenda for training interviewers in participating countries. While the content of this agenda is SHARE study specific, it might provide a useful basic template for other similar cross-national survey efforts. Organizations may add country-specific items to the model training agenda (e.g., tracing/locating steps that should be followed in their country and any relevant cultural considerations).

- Note that SHARE is a longitudinal (i.e., panel) study. However, new countries join at each wave, or a refresher sample is recruited – hence SHARE provides training for panel study and baseline study at most of its trainings.
- The model training agenda assumes that interviewers have already received basic training in General Interviewing Techniques (GIT). However, since SHARE wants to make sure that certain specific GIT interviewing conventions are always implemented, SHARE spends part of the study-specific training reviewing those.
- See the SHARE website for details about the study [60].

SHARE Model Training Agenda (na=not applicable)

Topic	**Purpose**	**Panel: Time (minutes)**	**Baseline: Time (minutes)**
Introductions, Welcome, and Logistics	Set the stage for this intense training.	15	30
SHARE Project and Questionnaire Overview	Explain the goals of the project and the importance of baseline and longitudinal sample.	45	45
Sample Overview	Understand how the sample was selected, sample eligibility, and response rate requirements.	30	60
GIT Requirements	Cover minimal GIT requirements, including when and how to contact sample, probes, feedback, etc.	60	60
Overview of the Sample Management System	Learn how to operate the SHARE electronic sample management system, assign result codes, and enter call notes. Introduce noncontact mock scenarios and test results.	60	90
Longitudinal Sample Management System	Introduce splitters, deceased, new eligible respondents, and additional result codes.	30	Na
Proxy Interviews	Explain how to identify and interview proxy respondents.	30	45
Nursing Homes	Explain how to contact respondents in nursing homes and to work with gatekeepers / potential proxy respondents.	30	Na

Topic	**Purpose**	**Panel: Time (minutes)**	**Baseline: Time (minutes)**
Overview of the Blaise Program	Explain the Blaise program conventions, including different types of questions, question wording, data entry, interviewer instructions, etc.	45	45
SHARE Questionnaire Walk-Through	Describe SHARE modules. Conduct a scripted review of the questionnaire, including spawning of additional line. Address main questions and issues that arise with different sections. Longitudinal: Describe longitudinal differences. Explain preloads. Address different questions arising from reinterviews.	330	240
End-of-Life Interviews (EOL)	Cover the concept of the EOL interview, approaching respondents, and administering the interview. Explain how to record these in the Sample Management System (SMS).	30	Na
Drop Off	Describe drop-off and the procedure for identifying and labeling drop-off appropriately. Explain procedure for administering drop-off and how to record these in SMS.	45	45
Physical Measurements; "Certification"	Have each interviewer demonstrate the ability to conduct physical measures..	30	60
Response Rates and Contact Efforts	Explain the importance of response rates and the reiteration of required contact effort per line. (Longitudinal: review only) Longitudinal: Cover panel care and effort requirements, including tracking effort.	45	90
Gaining Respondent Cooperation	Review eight concerns that interviewers are likely to encounter. Practice quick answers to several concerns. Note that longitudinal sample is more likely to encounter different types of resistance.	90	90
Practicing Household Introductions	Have interviewers team up in groups of 10 or so and each take a turn introducing the study.	Optional	60
Pair-wise Questionnaire Walk-through	This is an opportunity for interviewers to go through the questionnaire with a fellow interviewer. Use an abbreviated script. Switch at the half-point mark and complete the interview.	90	130
Pair-wise EOL Interview	Practice administering the End-of-Life interview.	45	Na
Administrative Wrap-Up	Answer outstanding questions.	30	30

- Total Time Training for the Panel Model: 1080 minutes (18 hours, 0 minutes)
- Total Time Training for the Baseline Model: 1120 minutes (18 hours, 40 minutes)

Appendix D

Example of an Interview Certification Form

NOTE:

- The aim of certification is to assess the interviewer's conduct of the interview, including introducing the study, doing the interview itself, using all appropriate respondent materials and interviewer aids, closing out the interview, and recording all required information in the sample management system.
- Specific studies should modify items in this form, as needed, to ensure that all key elements are measured in the certification.
    - For example, the template below assumes that an electronic sample management system (SMS) is used; if a paper coversheet is used to manage the sample, one should develop items appropriate for that system.
    - Similarly, the template assumes that the interview is programmed in Blaise; if data is collected via paper and pencil, one should check the interviewer's comfort in following routing instructions, choosing appropriate fills, etc.
- Additional items may be included on the form and scoring may be changed to suit the situation. Some potential additions might include (a) professionalism (e.g., pace, tone and emphasis of speech), (b) establishing rapport with the respondent, (c) introducing the study to the respondent, and (d) the administration of specific areas in the instrument, such as cognitive tests or mental health questions.

<table>
<tr><th colspan="4">Certifier Notes for Individual Certifications</th></tr>
<tr><td colspan="2">Interviewer:</td><td></td><td>Certifier:</td></tr>
<tr><td colspan="2">Time:</td><td></td><td>Location:</td></tr>
<tr><td colspan="4">CERTIFIER INSTRUCTIONS: Score each item 0, 1, or 2. 0 = Inadequate performance; 1 = Needs Improvement; 2 = Met Expectations. Use the Errors column to tally the number of times the interviewer makes general interviewing technique (GIT) errors in reading, probing, feedback, or clarification. Note question numbers of errors when possible.</td></tr>
<tr><th>Interviewing Skill</th><th>Score</th><th>Errors</th><th>Comments</th></tr>
<tr><td>On time and prepared for certification</td><td></td><td></td><td>Sample Management System running and ready to interview; for face-to face interview, have respondent materials ready, including copy of letter and brochure.</td></tr>
</table>

Correctly completing household listing/enumeration and screener			Make sure that the interviewer has completed the household listing/enumeration correctly; if not, tell him/her how to correct and proceed. **The interviewer will have to re-certify on the screener portion if this happens.**
Use of GIT probes and clarification			Should use standard GIT protocol as indicated; 1 - 2 errors - score 1; 3+ errors - score 0.
Use of neutral feedback			Interviewer should provide feedback for at least 30% of responses. Non-standard feedback counts as an error.
Verbatim question reading			Include pronunciation and emphasis in evaluation; 1-3 errors - score 1; 4+ errors - score 0.
Data Entry			General comfort with navigating in Blaise.
Post-interview process & contact person information			Interviewer should confirm all contact information for respondent and enter information for required number of contact persons.
Contact attempt record			Interviewer should enter a final contact attempt note which you will check before scoring. If 1 or 2 items are missing - score 1. If more than 2 items are missing - score 0.
TOTAL SCORE	0		

Total possible = 16 Certified = 12or higher Re-Certify = 10-11
Administrative Review will be required if score is less than 10.

GENERAL COMMENTS: Provide specific examples and question numbers of problem areas when possible. Note the way in which the interviewer administered the informed consent and reads the script to explain the need for obtaining information for contact persons.

Debriefing with Interviewer by [NAME]:	Date:

Notes: Include summary of recertification plan and retraining or practice interviews needed. Make note of areas that need close review on taped interviews.

Appendix E

Example of a Certification Checklist for Physical Measurements

NOTE:

- Interviewers should act as though this is a real interview. It is recommended that the person performing the certification ("certifier") observe a pair of interviewers where one acts as the respondent and the other is the interviewer being evaluated. If the "interviewer" asks questions during the certification, such as "should I ask/do this...", neither the certifier nor the "respondent" should respond.
- Make sure that the list of supplies is first checked off (interviewer has all materials ready or not).
- Observe that the interviewer reads all instructions and explanations to the respondent and enters the values correctly.
- If the interviewer performs a given activity correctly, make a "check" in the column labeled "Correct" for that activity; if the interviewer does not perform the activity correctly, circle the number in the column labeled "Incorrect" for that activity.
    - The numbers in the "Incorrect" column indicate the importance of the activity as defined by the researcher. Individual researchers can establish the relative "weight" of the error score, as necessary.
- For each physical measurement total the circled numbers and enter the sum in the row labeled "Total Incorrect;" also enter the total on the "Total Incorrect" line below the table. Assess whether or not the interviewer has passed the section. To be certified, the interviewer must successfully pass all sections.
        - Say, for example, the interviewer failed to correctly perform the Blood Pressure activities "arm on table" and "use correct cuff size." The certifier would circle the Incorrect scores of 2 and 4 respectively, for a Total Incorrect score of 6. Since the "Max incorrect to pass" is 3, the interviewer would not pass this section and would need to be re-trained and re-certified.
- At the end, be prepared to provide feedback regarding the certification items and whether the interviewer passed certification. Make a decision about whether to permit recertification (for all measures or for only those that the interviewer did not pass) and be sure to let some time pass before attempting a recertification.
- Retain final, signed records as documentation of this certification. Some large cross-national studies may require some form of documentation on interviewer certification levels across member countries.

PHYSICAL MEASUREMENTS CERTIFICATION CHECKLIST

Interviewer's Name: ___________________________

Certifier's Name: ______________________________

Date of Certification: ____________________________

Blood Pressure

Activity	Correct	Incorrect
Feet flat on floor/legs uncrossed		**2**
No smoking		**2**
Loose clothing/ no more than one layer		**4**
Arm on the table (or supported) at heart level		**2**
Use of correct cuff size		**4**
Tube of cuff hanging at inner crease of arm		**4**
Start at 180 SBP (Systolic Blood Pressure)		**2**
Re-inflation no sooner than 30-45 sec.		**4**
Re-inflation to first SBP + 20		**1**
Total Incorrect:		

Total incorrect Blood Pressure: ___________

Max incorrect to pass: 3 (4 or more needs re-certification)

Height

Activity	Correct	Incorrect
Shoes off		**4**
Heels to wall		**2**
Place sticky properly on wall		**2**
Orange triangular ruler on top of head, parallel to floor (fat edge against the wall)		**4**
Place metal tape measure properly and straight for accuracy in measuring height		**4**
Remove sticky from wall when done		**1**
For leg length, ask respondent to locate bony prominence and hold metal tape in place there; keep tape straight		**4**
Total Incorrect:		

Total incorrect (Height): ___________

Max incorrect to pass: 3 (4 or more needs re-certification)

Weight

Activity	Correct	Incorrect
Place scale on firm floor		**4**
Shoes off		**4**
Remove bulky clothes		**2**
Tap red label on scale; wait for “000.0”		**4**
Total Incorrect:		

Total incorrect (Weight): ___________

Max incorrect to pass: 2 (3 or more needs re-certification)

Waist

Activity	**Correct**	**Incorrect**
Ask respondent to identify umbilicus (navel) and hold cloth tape in place there		**3**
One layer of clothing		**3**
Tape snug but not tight		**1**
Check that tape is horizontal all around the R		**3**
Ask respondent to take normal breath and exhale, holding breath at end of exhalation		**1**
Record to nearest *centimeter*		**3**
Total incorrect:		

Total incorrect (Waist): ___________

Max incorrect to pass: 2 (3 or more needs re-certification)

Hip

Activity	**Correct**	**Incorrect**
Take measurement at respondent’s side		**2**
Place cloth tape at level of maximal protrusion of gluteal muscles		**3**
Tape snug but not tight		**1**
Check that it is horizontal all around the respondent		**3**
Move tape up and down to make sure measurement is taken at greatest diameter		**1**
Ask the respondent to take normal breath and exhale, holding breath at end of exhalation		**1**
Record to nearest *centimeter*		**3**
Total incorrect:		

Total incorrect (Hip): __________
Max incorrect to pass: 2 (3 or more needs re-certification)

Glossary

Accuracy	The degree of closeness an estimate has to the true value.
Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Adaptive behavior	Interviewer behavior that is tailored to the actual situation encountered.
Auxiliary data	Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in imputation techniques.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Bridge language	A language, common to both interviewers and respondents, that is used for data collection but may not be the first language of either person.
Certification	Objective assessment of performance. Based on pre-established criteria, the interviewer either meets the requirements and may proceed to conduct the study interview or does not meet the requirements and may either be permitted to try again or be dismissed from the study. Certification outcome should be documented and filed at the data collection agency.

Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: *Do you smoke?* *Yes* ___ *No* ___
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Coding	Translating nonnumeric data into numeric fields.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Computer assisted personal interviewing (CAPI)	A face-to-face interviewing mode in which a computer displays the questions onscreen, the interviewer reads them to the respondent, and enters the respondent's answers directly into the computer.
Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Contact attempt record	A written record of the time and outcome of each contact attempt to a sample unit.

Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Conversational interviewing	Interviewing style in which interviewers read questions as they are worded but are allowed to use their own words to clarify the meaning of the questions.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Coverage error	Survey error (variance and bias) that is introduced when there is not a one-to-one correspondence between frame and target population units. Some units in the target population are not included on the sampling frame (undercoverage), some units on the sampling frame are not members of the target population (out-of-scope), more than one unit on the sampling frame corresponds to the same target population unit (overcoverage), and one sampling frame unit corresponds to more than one target population unit.
Coversheet	Electronic or printed materials associated with each element that identify information about the element, e.g., the sample address, the unique identification number associated with an element, and the interviewer to whom an element is assigned. The coversheet often also contains an introduction to the study, instructions on how to screen sample members and randomly select the respondent, and space to record the date, time, outcome, and notes for every contact attempt.
Disclosure analysis and avoidance	The process of identifying and protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. For any proposed release of tabulations or microdata, the level of risk of disclosure should be evaluated.

Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Ethics review committee or human subjects review board	A group or committee that is given the responsibility by an institution to review that institution's research projects involving human subjects. The primary purpose of the review is to assure the protection of the safety, rights and welfare of the human subjects.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Hours per interview (HPI)	A measure of study efficiency, calculated as the total number of interviewer hours spent during production (including travel, reluctance handling, listing, completing an interview, and other administrative tasks) divided by the total number of interviews.
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Interpenetrated sample assignment, interpenetration	Randomized assignment of interviewers to subsamples of respondents in order to measure correlated response variance, arising from the fact that response errors of persons interviewed by the same interviewer may be correlated. Interpenetration allows researchers to disentangle the effects interviewers have on respondents from the true differences between respondents.

Interviewer design effect ($Deff_{int}$)	The extent to which interviewer variance increases the variance of the sample mean of a simple random sample.
Interviewer effect	Measurement error, both systematic and variable, for which interviewers are responsible.
Interviewer variance	That component of overall variability in survey estimates that can be accounted for by the interviewers.
Item nonresponse, item missing data	The absence of information on individual data items for a sample element where other data items were successfully obtained.
Longitudinal study	A study where elements are repeatedly measured over time.
Majority country	A country with low per capita income (the majority of countries).
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Microdata	Nonaggregated data that concern individual records for sampled units, such as households, respondents, organizations, administrators, schools, classrooms, students, etc. Microdata may come from auxiliary sources (e.g., census or geographical data) as well as surveys. They are contrasted with macrodata, such as variable means and frequencies, gained through the aggregation of microdata.
Mode	Method of data collection.
Noncontact	Sampling units that were potentially eligible but could not be reached

Nonresponse bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.
Nonresponse error	Survey error (variance and bias) that is introduced when not all sample members participate in the survey (unit nonresponse) or not all survey items are answered (item nonreponse) by a sample element.
Open-ended question	A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers. Example: *What is your occupation?* *Please write in the name or title of your occupation*___________
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Prescribed behaviors	Interviewer behaviors that must be carried out exactly as specified.
Processing error	Survey error (variance and bias) that arise during the steps between collecting information from the respondent and having the value used in estimation. Processing errors include all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in the implementation of systems and estimates. In survey data, processing errors may include errors of transcription, errors of coding, errors of data entry, errors in the assignment of weights, errors in disclosure avoidance, and errors of arithmetic in tabulation.
Proxy interview	An interview with someone (e.g., parent, spouse) other than the person about whom information is being sought. There should be a set of rules specific to each survey that define who can serve as a proxy respondent.

Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Reliability	The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Sample element	A selected unit of the target population that may be eligible or ineligible.

Sample management system	A computerized and/or paper-based system used to assign and monitor sample units and record documentation for sample records (e.g., time and outcome of each contact attempt).
Sampling error	Survey error (variance and bias) due to observing a sample of the population rather than the entire population.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Simple random sampling (SRS)	A procedure where a sample of size n is drawn from a population of size N in such a way that every possible sample of size n has the same probability of being selected.
Standardized interviewing technique	An interviewing technique in which interviewers are trained to read every question exactly as worded, abstain from interpreting questions or responses, and do not offer much clarification.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.

Tailoring	The practice of adapting interviewer behavior to the respondent's expressed concerns and other cues, in order to provide feedback to the respondent that addresses his or her perceived reasons for not wanting to participate.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Unique Identification Number	A unique number that identifies an element (e.g. serial number). That number sticks to the element through the whole survey lifecycle and is published with the public dataset. It does not contain any information.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Validity	The extent to which a variable measures what it intends to measure.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

References

[1] Alcser, K. H., & Benson, G. D. (2005). The SHARE train-the-trainer program. Ch. 6 In A. Boersch-Supan & H. Juerges (Eds.), *The survey of health, ageing and retirement in Europe – Methodology*. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging.

[2] Alcser, K. H., & Benson, G. D. (2008). Training for SHARE Wave 2. Ch. 8.3 in A. Boersch-Supan, A. Brugiavini, H. Juerges, et al. (Eds.). *Health, Ageing and Retirement in Europe (2004-2007)*. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging.

[3] Axinn, W. G. (1989). Interviewers and data quality in a less developed setting. *Journal of Official Statistics*, *5*(3), 265-280.

[4] Axinn, W.G., Fricke, T.E., & Thornton, A. (1991). The micro-demographic community study approach: Improving survey data by integrating the ethnographic method. *Sociological Methods Research*, *20*(2), 187-217.

[5] Beatty, P. (1995). Understanding the Standardized/Non-Standardized Interviewing Controversy. *Journal of Official Statistics, 11*(2), 147-160.

[6] Belli, R., Shay, W., & Stafford, F. (2001). Event history calendars and question lists. *Public Opinion Quarterly, 65*(1), 45-74.

[7] Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.

[8] Billiet, J., & Loosveldt, G. (1988). Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly*, *52*(5), 190-211.

[9] Börsch-Supan, A., Jürges, H., & Lipps, O. (2003). *SHARE: Building a panel survey on health, aging and retirement in Europe*. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging.

[10] California Health Interview Survey (CHIS). (2007). *CHIS 2005 Methodology series: Report 2 — Data collection methods*. Los Angeles, CA: UCLA Center for Health Policy Research.

[11] Cannel, C. F. et al. (1977). *A summary of studies of interviewing methodology. Vital and Health Statistics: Series 2, Data Evaluation and Methods Research* (No. 69. DHEW Publication No. (HRA) 77-1343). Washington, DC: U.S. Government Printing Office. Retrieved December 29, 2009, from http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDeta

ils/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED137418&ERICExtSearch_SearchType_0=no&accno=ED137418

[12] Cantave, M., Kreuter, F., & Alldredge, E. (2009). *Effect of pay structure on interviewer productivity.* Poster presented at the 2009 AAPOR meetings.

[13] Christopher, S., McCormick, A. K. H. G., Smith, A., & Christopher, J. C. (2005). Development of an interviewer training manual for a cervical health project on the Apsáalooke reservation. *Health Promotion Practice*, *6*(4), 414-422.

[14] Conrad, F., & Schober, M. F. (1999). Conversational interviewing and data quality. *Proceedings on the federal committee on statistical methodological research conference*. Retrieved February 14, 2008, from http://www.bls.gov/ore/abstract/st/st990250.htm

[15] Davis, D. W. (1997). Nonrandom measurement error and race of interviewer effects among African Americans. *Public Opinion Quarterly, 61*(1), 183-207.

[16] Doyle, J. K. (2004). Introduction to interviewing techniques. Ch. 11 in Wood. D. W. (Ed.), *Handbook for IQP advisors and students* (Prepared for the Interdisciplinary and Global Studies Division). Worcester, MA: Worcester Polytechnic Institute. Retrieved March 26, 2010, from http://www.wpi.edu/academics/GPP/Students/handbook.html

[17] European Social Survey. (2004). *Field procedures in the European Social Survey: Enhancing response rates*. Retrieved January 24, 2008, from http://www.europeansocialsurvey.org/index.php?option=com_docman&task=doc_view&gid=160

[18] Federal Committee on Statistical Methodology. (1998). Interviewer training. Ch. 5 in *Training for the future: Addressing tomorrow's survey tasks* (pp. 50-61). (Report number NTIS PB99-102576). Washington, DC: U.S. Government Printing Office. Retrieved March 26, 2010, from http://www.fcsm.gov/working-papers/1998_intro-appD.html

[19] Fowler, F. J. Jr. & Mangione, T.W. (1985). The value of interviewer training and supervision (Final report to the National Center for Health services Research). Boston, MA: Center for Survey Research.

[20] Fowler, F. J. Jr. & Managione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Beverly Hills, CA: Sage Publications.

[21] Galbraith, M. (2003). *Adult learning methods* (3rd ed.). Malabar, FL: Kreiger Publishing Co.

[22] Glewwe, P. (2005). Chapter IV: Overview of the implementation of household surveys. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), *Household surveys in developing and transition countries*. New York, NY: United Nations. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[23] Groves, R. (1989). *Survey errors and survey costs*. New York, NY: John Wiley & Sons.

[24] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[25] Groves, R. M., & McGonagle, K. A. (2001). A theory-guided interviewer training protocol regarding survey participation. *Journal of Official Statistics*, *17*(2), 249-265.

[26] Hansen, M. H., Hurwitz, W. N., Marks, E. S., & Mauldin, W. P. (1951). Response errors in surveys. *Journal of the American Statistical Association, 46*(254), 147-190.

[27] Hatchett, S., & Schuman, H. (1975). White respondents and race-of-interviewer effects. *Public Opinion Quarterly*, *39*(4), 523-528.

[28] Hursh-César, R. P. (1976). *Third world surveys: Survey research in developing nations*. Delhi: Macmillan Company of India.

[29] Jansen, H. A. F. M., Watts, C., Ellsberg, C., Heise, L., & García-Moreno, C. (2004). Interviewer training in the WHO multi-country study on women's health and domestic violence. *Violence against Women*, *10*(7), 831-849.

[30] Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57*(1), 1-28.

[31] Kessler, R. C., Ustun, T. B., & World Health Organization. (2008). *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge; New York :Geneva: Cambridge University Press; Published in collaboration with the World Health Organization.

[32] Kish, L. (1962) Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, *57*(297), 92-115.

[33] Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision* (2nd ed.). Thousand Oaks, CA: Sage Publications.

[34] Lee, R.M. (1993). Doing research on sensitive topics. London, UK: Sage

Publications.

[35] Liamputtong, P. (2010). *Performing qualitative cross-cultural research*. Cambridge, UK: Cambridge University Press.

[36] Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

[37] Martin, E. (1996). Household attachment and survey coverage. *Proceedings of the section on survey research methods* (pp. 526-531). Alexandria, VA: American Statistical Association.

[38] Nyandieka, L. N., Bowden, A., Wanjau, J., & Fox-Rushby, J. A. (2002). Managing a household survey: A practical example from the KENQOL survey. *Health Policy and Planning*, *17*(2), 207-212.

[39] O'Brien, E. M., Mayer, T. S., Groves, R. M., & O'Neill, G. E. (2002). Interviewer training to increase survey participation. *Proceedings of the section on survey research methods*. Alexandria, VA: American Statistical Association.

[40] Office of Management and Budget. (2006). *Standards and guidelines for statistical surveys*. Washington, DC: Office of Information and Regulatory Affairs, OMB. Retrieved January 10, 2010, from http://www.whitehouse.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf

[41] O'Muircheartaigh, C. & Marckward, (1980). An assessment of the reliability of World Fertility Study Data. *Proceedings of the World Fertility Survey conference, Vol. 3,* 305-379. The Hague, The Netherlands: International Statistical Institute.

[42] Papanek, H. (1979). Research on women by women: Interviewer selection and training in Indonesia. *Studies in Family Planning*, *10*(11/12), 412-415.

[43] Pennell, B-E., Harkness, J., & Mohler, P. (2006). *Day 8: Data collection, interviewers, and interviewer training*. *SRC Summer Institute Presentation*. University of Michigan, Ann Arbor, MI.

[44] Pennell, B-E., Harkness, J., Levenstein, R., & Quaglia, M. (2010). Challenges in cross-national data collection. In J. Harkness, B. Edwards, M. Braun, T. Johnson, L. E. Lyberg, P. Mohler, B-E. Pennell, & T. Smith. (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts*. Hoboken, NJ: John Wiley & Sons.

[45] Pennell, B-E., Mneimneh, Z., Bowers, A., Chardoul, S., Wells, J. E., Viana, M. C., et al. (2009). Implementation of the World Mental Health survey initiative. In R. C. Kessler & T. B. Ustun (Eds.), *Volume 1: Patterns of mental illness in the WMH surveys*. Cambridge, MA: Cambridge University Press.

[46] Quaglia, M. (2006). *Interviewers and interpreters: Training for a biographical survey in Sub-Saharan Africa*. Paper presented at Proceedings of Q2006 European Conference on Quality in Survey Statistics. Retrieved April 30, 2008, from http://www.statistics.gov.uk/events/q2006/agenda.asp

[47] Schaeffer, N. C. (1980). Evaluating race-of-interviewer effects in a national survey. *Sociological Methods & Research*, *8*(4), 400-419.

[48] Schnell, R. & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics, 21*(3)*,* 389–410.

[49] Schuman, H., & Converse, J. M. (1971). The effects of black and white interviewers on black responses in 1968. *Public Opinion Quarterly*, *35*(1), 44-68.

[50] Stycos, J. M. (1952). Interviewer training in another culture. *Public Opinion Quarterly, 16*(2), 236-246.

[51] Suchman, S. & Jordan, B. (1990). Interactional troubles in face to face survey interviews. *Journal of the American Statistical Association, 85*, 232-241.

[52] Sudman, S. (1966). New approaches to control of interviewing costs. *Journal of Marketing Research*, *3*(1), 55-61.

[53] Ustun, T. B., Chatterji, S., Mechbal, A., & Murray, C. J. L. (2005). Chapter X: Quality assurance in surveys: Standards, guidelines, and procedures. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), *Household surveys in developing and transition countries*. New York, NY: United Nations. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[54] Vaessen, M., Thiam, M., & Le, T. (2005). Chapter XXII: The demographic and health surveys. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), *Household surveys in developing and transition countries*. New York, NY: United Nations. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[55] Weisberg, H. (2005). *The total survey error approach*. Chicago, IL: University of Chicago Press.

[56] West, B. & Olson, K. (2009) How much of interviewer variance is really nonresponse error variance? *Michigan Program in Survey*

Methodology Brownbag Presentation. University of Michigan, Ann Arbor, MI.

Internet Links

[57] Afrobarometer Survey. Retrieved July 1, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[58] Asian Barometer. Retrieved July 1, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

[59] European Social Survey. Retrieved July 1, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=23&Itemid=318

[60] Survey of Health, Ageing and Retirement in Europe. Retrieved July 1, 2010, from http://www.share-project.org

XI. Pretesting

Rachel Caspar and Emilia Peytcheva

Introduction

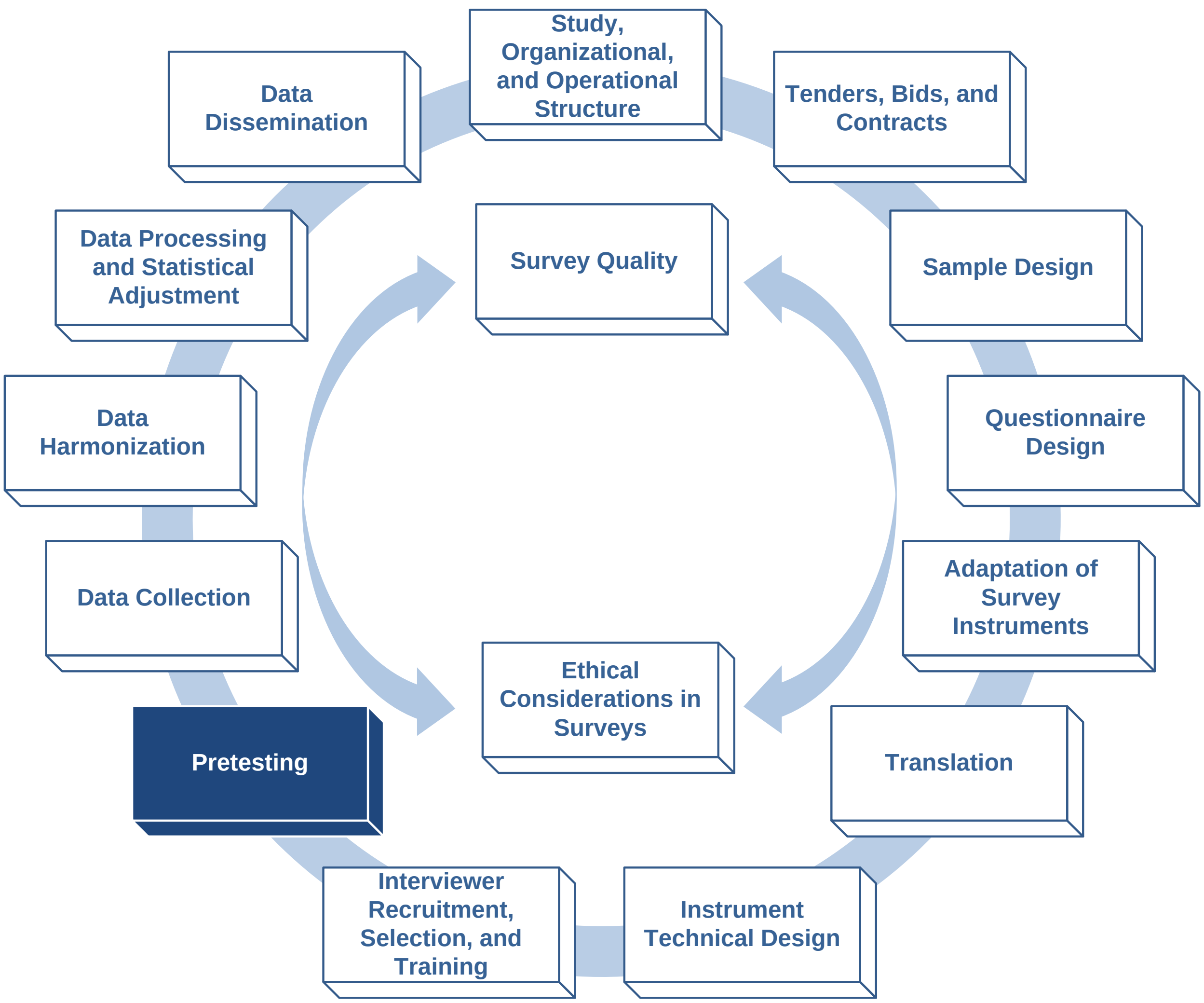
Pretesting involves a series of activities designed to evaluate a survey instrument's capacity to collect the desired data, the capabilities of the selected mode of data collection, and the overall adequacy of the field procedures. Throughout this text we refer to a "pretest" as the collection of all those techniques and activities that allow researchers to evaluate survey questions and survey procedures before data collection begins. In contrast, we use the term "pilot study" to refer to pretesting procedures that employ all the procedures and materials involved in data collection (regardless of how small of a scale) before the actual data collection begins. Pilot studies are also referred to as "dress rehearsals", or "field tests" in the survey literature and they have a specific goal – from estimating response rates under a particular recruitment protocol to identifying an optimal design characteristic (e.g., incentive amount) through experimentation.

This chapter provides examples mainly based on U.S. surveys that sample ethnic minorities and immigrants and are administered in different languages, but attempts to extrapolate experiences and lessons learned to cross-national surveys. Table 1 is a summary of the most commonly used pretesting techniques, such as pilot studies, concurrent or retrospective think aloud techniques, focus groups, and behavior coding.

When multiple languages are used in the same survey, pretesting the different language versions is an essential part of ensuring measurement equivalence and testing translations with the target population (see Translation). In addition, it is often difficult to employ the same mode of data collection across countries participating in a cross-national project. It is important to test in advance the suitability of the selected mode for the survey topic and population (see Data Collection). Pretesting techniques may have limited application in a given context and culture. Research into how pretesting strategies may need to be tailored to suit different populations is only beginning to be undertaken systematically.

Figure 1 shows pretesting within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To ensure that the versions of the survey instrument adequately convey the intended research questions, measure the intended attitudes, values, reported facts and behaviors, and that the collections of data are conducted according to specified study protocols in every country and in every language.

1. **Identify what the pretest should achieve and choose a pretest design that best fits the study goals.**

Rationale

To make the best use of the various pretesting techniques, it should be determined in advance what issues have to be addressed — whether the researchers want to test all field procedures, or only the survey instrument (or parts of it), or the equivalence of the survey instrument across languages and modes of data collection. Pretesting should be done in each country participating in the research. It should be noted that available pretesting techniques may vary across countries, depending on testing traditions, resources, and trained staff. Even if some or all of the questions have been used in other studies, pretesting for the local context is necessary in order to assess their performance in the mode and question order in the current study, and with the target population.

Procedural steps

- Using Table 1 as an aid, decide what type of pretesting technique(s) will best fit the study's purpose and the culture within which the study will be conducted.
  - Table 1 presents the most commonly used pretesting techniques.
  - It provides a brief description, strength and weaknesses, and context in which a particular technique is typically used.

- Assess whether to conduct the pretest(s) in-house or to contract the testing to an outside organization.

- Establish a time schedule that adequately matches the pretesting design, allowing sufficient time to implement any revisions which may be deemed necessary prior to implementing the full study.

- Budget accordingly. Plan also on expenses related to interviewer and staff training, respondent recruitment, and incentives.

Lessons learned

- When conducting cross-national cognitive interviewing projects, care must be taken to ensure that comparable procedures are utilized across all countries involved. Even when the same pretesting technique is used, if its implementation varies drastically across countries, it becomes impossible to determine whether observed differences are due to differences in the response process, translation, or the conceptual spectrum. For example, it is not safe to assume that procedures for conducting cognitive interviews will be the same across all countries. Differences may exist, for example, in the experience of the interviewers, the location of the interviewing, methods used to recruit participants, and approaches to creating the interviewing protocol. Recent work in seven countries (eight languages) has focused on creating a common approach to cognitive interviewing for questions designed to measure health status [18]. To ensure equivalence, all parties involved in the project agreed upon the method to be used for recruiting participants, administering the protocol, and documenting results.

- It is important to note that even when standardized protocols are used across countries, pretesting techniques such as cognitive interviews do not work equally well across cultural groups without modification [9] [20]. Pan [19] investigated the efficacy of concurrent think aloud as a pretesting strategy with Chinese respondents. Her investigation identifies challenges and limitations of taking methods developed in one language and culture and directly applying them to another, and points to the need to include consideration of sociolinguistic conventions appropriate to different cultural groups when conducting cognitive interviews.

- While focus groups are a quick way to gain in-depth insight into participant knowledge and attitudes, Helitzer-Allen et al [14] argue that studies, particularly in the health field, are relying too heavily on this method. While previous research has shown that focus groups are generally useful in collecting sensitive matter, some topics are exceptions. In a case study in Malawi, adolescent girls were interviewed using two different methods: in-depth interviews and focus group discussions. The study, conducted through the National AIDS Control Programme, utilized mixed methods through quantitative data collection of census information and highly-structured questionnaires as well as qualitative observation, less-structured interviews, and focus groups. Overall, the study found that studies cannot solely rely on focus groups due to the sensitivity of some material being prevalent to the degree that individuals will not discuss the material in front of one another. For the female subjects in Malawi, menstruation was too sensitive to discuss in focus groups. The authors recommend that researchers use both methods, with in-depth interviews conducted before focus groups. They found that by asking the females sensitive questions during their in-depth interviews, they were then able to follow up some of the interview questions by asking if the subject would be willing to discuss this topic in groups of numerous girls.

- Culture or language specific probes may be needed to best test the translation/adaptation of a survey instrument. For example, the Census Bureau conducted cognitive tests of the translations of the introductory letters and informational brochures for the American Community Survey in seven languages [21]. The focus of the study was to examine how cognitive interviews work in non-English languages given cultural differences in communication. Remarkable differences in the way participants from different language groups provided responses were reported. For example, the Chinese and Korean respondents tended to provide limited responses and their answers were not focused on the topic, while Russian respondents showed a tendency to always give 'confident' answers. Such differences in response patterns raise questions related to data quality and comparability of cognitive interview results across language groups and stresses the importance of culturally appropriate modifications to existing pretesting techniques.

- A related practical question is whether to create cognitive protocols in English and then translate into the target languages, or develop the protocols directly into the target languages, accounting for different cultural norms and socialization styles. Each approach has benefits and weaknesses that have to be weighed against one another given the specific survey conditions (e.g., simultaneous development of the protocol guides may not be as feasible in multilingual projects as it is in bilingual studies). Goerman and Caspar [10] discuss approaches for creating protocol guides in multiple languages that ensure culture and language appropriateness and present strategies for respondent recruitment, interviewer selection and training that allow adequate testing of instrument translation.

2. Combine pretesting techniques to create a comprehensive design plan that takes advantage of the strengths and minimizes the weaknesses of each method.

Rationale

Pretesting techniques often supplement one another and can logically be combined to maximize the efficiency of the pretest design (see Table 1). For example, to minimize cost, one can consider pretesting a questionnaire using an expert review. Once the questionnaire is revised based on reviewers' comments, participants for cognitive interviews can be recruited, or a pilot study can be launched. It is important to take cultural norms and traditions, as well as interviewer characteristics (see Data Collection), into account when choosing pretesting methods. The most appropriate combinations of pretesting techniques may vary across countries involved in the study. This should be taken into account when results from the different pretests are evaluated and compared.

Procedural steps

- Begin with pretesting methods that focus on specific aspects of the study (for example, wording of particular questionnaire items, comprehensibility of the informed consent, procedures for interviewers to follow in administering the survey) before moving to techniques that pull all aspects of the project into a more comprehensive study.
  - For example, consider a focus group or in-depth interviews for initial questionnaire construction, cognitive interviews for questionnaire development and refinement, and a field pilot study for an overall test of the survey instrument and field procedures. Often, a pilot study with robust samples can be the best way to test the survey instrument as data analyses with sufficient power can be the most effective way to ascertain if the questionnaire is working as intended.

- Discuss every round of changes introduced to the questionnaire with the coordinating center and test again—consider several iterations of testing, rather than one large scale pretest (also see Translation).

Lessons learned

- In preparation for the shift from a paper-and-pencil instrument to a computer-assisted instrument incorporating a large audio computer-assisted self-interview (ACASI) component, the U.S. Substance Abuse and Mental Health Services Administration (SAMHSA) implemented a comprehensive pretesting plan [8]. The overarching goal of the pretesting was to develop an optimal computerized instrument on the sensitive topic of drug usage. It was also essential that any differences in reporting due to the mode change be identified so that data users would understand how to interpret trend lines from the data. Pretesting work first concentrated on small-scale cognitive laboratory testing to determine the best way to structure the instrument, to train respondents to use the computer for the ACASI components, to determine the voice to be used for the audio component, and to assess respondents' ability to enter different types of data into the computer (e.g., open-ended responses). Based on results from these laboratory studies, a pilot study was conducted to evaluate interviewer training materials and to collect sufficient data to determine how the mode change impacted reporting. After changes were made based on this field pilot study, a larger pilot study, incorporating an experimental design, was conducted. Finally, the revised instrument and procedures were implemented in a split-sample comparison with the original paper-and-pencil instrument during data collection to allow researchers to assess the impact on the trend lines.

3. Train or hire staff members who are able to adequately implement the chosen pretesting technique(s).

Rationale

The selected pretesting procedures may require skills other than those that available interviewers possess. For example, cognitive interviewing requires a discursive interviewing style which is different from traditional standardized interviewing and requires additional training. Sufficient time and effort should be allowed to train staff members and develop protocols that correspond to the selected pretest design.

Procedural steps

- Establish standardized protocols across countries on:
  - How to select staff members for the pretest.
  - How to train staff members for the pretest.
  - How to best convey the objective of the task (technique dependent).
  - How to standardize or harmonize the pretesting protocol.
  - What methods will be used to monitor quality. Audio and video recordings are common during cognitive interviews and focus groups to help with the reporting process. However, such recordings can also be used to monitor interviewers and focus group moderators to ensure adherence to the pretesting protocol guide. Computer-Assisted Recorded Interviewing (CARI) allows for monitoring during field pretest and field data collection to detect interviewer fraud and ensure data quality. For a larger discussion on the importance of quality control and how to incorporate it at various survey stages see Survey Quality and Data Collection.
  - How results of the pretest will be analyzed (e.g., whether the analysis will be qualitative or quantitative).
  - How to decide on changes for the survey instrument.
- If different pretest designs are employed in different countries, select interviewers, training, and protocol that match the chosen technique; when the same techniques are used in various countries, harmonize all procedures.
- Consider interviewer characteristics as they may affect the outcome of a pretest in certain culture more than others (e.g., conversational styles in many cultures are largely determined by the status or education of the actors in the social hierarchy).
- Monitor interviewer behavior to ensure data quality.

Lessons learned

- Training interviewers and using structured protocols are important components of cognitive interviewing that can help reduce measurement error. In one project, after a series of cognitive interviews with Vietnamese respondents, a Vietnamese team member insisted that the respondents were not able to understand true/false items because they were not posed as questions. The translation of the items was shown to a Vietnamese co-worker who, however, did not find any problems. When the interview tapes were reviewed, it was discovered that the interviewer behaved inappropriately and read every item without stopping to get the respondent's answers. In some interviews, interviewers even read the skip instructions to the respondents.

4. **Conduct the pretest in the same mode of data collection (interviewer administered or self-administered) as the main survey.**

Rationale

Interviewer-administered questionnaires involve listening; self-administered questionnaires involve reading. Interviewer-administered questionnaires involve social interaction between the interviewer and the respondent; self-administered do not. Interviewer-administered questionnaires do not require the respondent to navigate and worry about skip patterns; self-administered do.

Interviewer-administered and self-administered questionnaires produce different context effects (e.g., recency and primacy) and provoke different need for socially desirable responding (see Data Collection). In order to determine how well proposed procedures will work in the field, not only pilot studies, but also cognitive laboratory pretests should be conducted in the same mode as the final survey.

Procedural steps

- If different modes of data collection are going to be employed across countries, consider pretesting in the respective modes.

- Some pretest techniques are not portable across modes (for example, behavior coding); others require modification. Adapt pretesting techniques to better match the mode of survey data collection (e.g., [22]).

- Use the latest version of the instrument and the respective materials (e.g., show cards, event history calendars).

- Use field administration procedures planned for production data collection.

Lessons learned

- Since each mode of data collection has its specific characteristics, it is important to pretest the survey instrument and procedures in every mode that will be used, even when the survey questionnaire is not translated to a different language. In fact, a change in mode may necessitate changes in wording or changes in design in order to achieve measurement equivalence. For example, cognitive testing for the 2001 US Census showed that more redundancy was needed in the instructions to the "respondent race" question for the respondents to be able to follow the "select one-or-more" option in telephone administration [6]. A slightly reworded version of the instructions and question stem resulted in better understanding of the intent of the question over the phone compared to what was needed when asking the question as it appeared in the mail questionnaire [17].

5. **Conduct the pretest with the same target population as the target population for the survey.**

Rationale

In order to most effectively pretest the survey instrument or field procedures, conduct the pretest with respondents from the intended target population or, as relevant, a sub-group within the target population. Similarly, the population of a pilot study should be an adequate reflection of the survey target population. For example, if the survey design involves oversampling of certain ethnic groups, the pretest sample should also include reasonable representation of these groups. A pretest with sample persons from the target population will most accurately reflect what will happen during real data collection in terms of cooperation, respondent performance, total interview length, questionnaire performance, survey costs, etc.

Procedural steps

For all pretesting techniques:

- Tailor subject or respondent recruitment to the population of interest.

- Prepare all necessary materials that would be used in the main survey, including an informed consent form that reflects the goals and risks of the pretest study (which may be different from the main survey).

- Select a sample size that is suitable for the chosen pretesting method.

- Monitor pretest participant recruitment to ensure best use of the chosen pretesting method.

For pilot studies:

- Select a sample large enough to provide sufficient statistical power to detect the differences that will answer the research questions identified in your pilot study analysis plan (with suitable allowances for nonresponse, eligibility, etc.).

- Follow the sample selection protocol planned for the final study.

- Monitor the sample selection

Lessons learned

- Select respondents from the survey target population; however, keep in mind that sometimes "survey-trained" respondents may be needed to detect potential problems. A study on pretesting by Hunt et al. [15] demonstrated that the general population may not be a good judge of the quality of survey questions, even when this is the target population. The researchers introduced obvious errors in the short questionnaire (e.g., missing response alternatives, inappropriate vocabulary) and asked respondents to be critical of the questions while answering them. Only a third of the sample noticed a missing response alternative; almost no one commented on "double-barreled" questions and "loaded" words. One possible explanation is that all of the respondents had roughly the same low level of survey experience.

6. Pretest the survey instrument or part of it in each country and in each language.

Rationale

When possible, the natural flow of the survey instrument should be tested for each culture and language in order to avoid awkward conversational situations, question order with unpredictable culture-dependent context effects, question repetition not intended in the source, or other culture-specific problems.

Procedural steps

- Select staff members who are fluent in the language of the pretest and sensitive to nuances.

- Standardize key components of training procedures across organizations.

- Develop a procedure for how problems will be reported and addressed in a consistent manner between testing sites.

Lessons learned

- Work conducted by the U.S. Census Bureau to develop a bilingual (English/Spanish) decennial census form has involved cognitive testing to identify potential problems with the layout of the form, to test respondents' ability to correctly navigate through the form, and to assess the quality of the Spanish translation [11]. Testing did not directly assess the English questions, as the wording of the English items had already been nearly finalized. As part of one particular study, cognitive interviews were conducted with monolingual Spanish speakers and bilingual Spanish-dominant speakers to focus on translation issues. Results from the testing indicated specific questions that were problematic for Spanish speakers. However, because there was no comparable group of English speakers included in the testing, it was difficult to determine whether the problems were confined to the translated items or would also be problematic for respondents who read the English wordings. To eliminate this problem, in a second round of testing, monolingual English respondents were included as well. The inclusion of these respondents allowed the researchers to identify where problems with the Spanish translation was due to specific choices made in the translation and where concepts were unclear for the Hispanic respondents as opposed to questions that were equally unclear for both English and Spanish speakers.

- Large established cross-cultural studies vary in the type and amount of pretesting they do.
  - Prior to the start of Round 1, the European Social Survey (ESS) source questionnaire was pretested using "interaction analysis" (i.e., behavior coding) to identify questions which were problematic for the interviewer or respondent. Problem questions were modified and the questionnaire was translated into various languages. In accordance with ESS Round 5 specifications, each participating country was required to pretest its translated questionnaire on a quota controlled, demographically balanced sample of around 50 people. The aims of pretesting were, at a minimum, to check routing and comprehension. Ideally the pretests could also be used to check for equivalence between the translated version of the questionnaire and the source. Countries were encouraged to tape record interviews, conduct respondent and/or interviewer

debriefings, and use cognitive interviewing to test for equivalence. The specifications note that these pretests occurred after the source questionnaire had been finalized and that opportunities to amend the source questionnaire were extremely limited at this point [29].

- The Survey of Health, Ageing and Retirement in Europe (SHARE) utilized a four-stage questionnaire development process. In the first stage, working groups produced an English-language draft questionnaire which drew from preexisting survey instruments. The draft questionnaire was piloted in the UK in September, 2002. Based on the lessons from this pilot, the English-language questionnaire was revised and translated into all of the SHARE languages. In the second stage, the translated questionnaires were simultaneously piloted in all SHARE countries, each testing a quota sample of 75 persons. In the third stage, after further revisions to the survey instrument, the full questionnaire was tested in all countries using probability samples (some 100 primary respondents per country plus their spouses). This all-country pretest also tested the country-specific logistics and the procedures to achieve probability samples. During the fourth stage, pilot and pretest results were statistically analyzed, leading to the final design of the questionnaire [30].

7. **Evaluate the results of the pretest.**

Rationale

The goal of the pretest is to identify problems in the questionnaire and study design in each country. The results of the pretest have to be evaluated to determine the best way to fix existing problems without introducing new ones. Changes to the survey instrument and design should be considered in the context of the whole study -- changes that fix a problem in one country may introduce a problem in another. The coordinating center should decide whether minor differences that still preserve the measurement equivalence of the survey instrument across countries can be tolerated (see Translation). Any introduced changes should also be pretested to avoid unforeseen errors (also see Instrument Technical Design).

Procedural steps

- Examine the findings of each pretesting technique used and identify the causes of the any problems discovered.
  - Decide in advance what constitutes a problem. For example, the 10%-20% rule is often used in behavior coding to flag questions: if a question is misread or misunderstood by 10%-20% of respondents, then it is considered problematic. It is important to

note that the appropriate threshold for any particular study is often determined from the distribution of coded errors (which is dependent on the coding scheme and instructions for code assignments).
    - Look for problems that are common across interviews, but also be aware that a problem may be important even if it occurred in only one interview. This is especially important when qualitative techniques are used – in order to determine what constitutes a problem, all possible factors that play a role in the pretest should be considered.
    - Examine in what situations and with what types of respondents problems occur.

- If a pilot study has been conducted:
    - Review response distributions and item nonresponse for key study variables.
    - Review interview length.
    - For attitudinal and value variables, check whether items group together as intended in the survey (e.g., perform confirmatory factor analysis).
    - Solicit and review feedback from interviewers.

- Report the results and proposed changes to the coordinating center. It is important that the timing and documentation of the pretest are coordinated across participating countries to allow overall comparison of results and propose meaningful changes.

- If changes are introduced to the questionnaire or design procedures, plan for another pretest.

Lessons learned

- Pretesting techniques and the results they yield are meaningful only when the selected procedures are culturally appropriate. Not many pretesting techniques have been tested and studied across countries; thus, some may not be successfully implemented and lead to meaningless results in certain cultures. For example, studies in psycholinguistics have demonstrated difference in cognitive tendencies among Chinese and English speakers to use counterfactual reasoning [3]. When asked what their thoughts would have been on a hypothetical legislation by their government, Hong Kong respondents consistently responded that the government has not proposed such legislation. Chinese speakers were less attuned to hypothetical thinking because their language does not mark counterfactuals differently from conditional statements. Such examples suggest that certain cognitive laboratory methods (for example, vignettes) may be of limited use in some cultures.

8. Fully document the pretesting protocol and findings.

Rationale

Providing a permanent record of problems encountered during the pretest(s) and any changes made to the questionnaire, respondent materials, and field procedures aids staff and researchers working on similar studies or on later rounds of the same study.

Procedural steps

In a manner consistent across countries:

- Document the pretest sample selection and recruitment method, including the sampling frame and sample size.
- Document the geographical location of the pretest.
- Document respondent characteristics.
- Documents mode(s) of pretest administration.
- Document dates of data collection and organization(s) conducting the interviews.
- Document types of staff conducting pretest (e.g., experienced interviewers, supervisors) and training.
- Document all materials used in the pretest.
- Describe pretest findings and their implications.
- Document any changes made to the survey instrument and the pretesting source that lead to these changes.
- Document the number and types of pretest.

Lessons learned

- The documentation can serve as a resource for future studies. For example, researchers within a U.S. Federal Interagency Group have developed Q-BANK, a database of questions for national health surveys maintained by their Questionnaire Design Research

Laboratory (QDRL) at the National Center for Health Statistics, Center for Disease Control (CDC). The database catalogues tested questions and links each question to cognitive testing findings. Questions are searchable not only by content or subject matter (e.g., asthma questions, cancer questions, demographics), but also by question type (e.g., objective characteristics, behavioral reports, attitudes), response category type (e.g., yes/no, open-ended, quantity), and response error type (e.g., problems with terms, recall problems). A statistical tool has been developed that performs basic statistical procedures on questions in the database.

Q-BANK, when completed, will centralize cognitive testing reports with links to specific questions and topic areas and will advance the field by: 1) serving as a resource in the development of new questions, 2) allowing question and response error comparisons across studies, 3) performing analysis on the characteristics of questions contributing to specific response errors, and 4) serving as a research tool investigating response error.

Q-BANK is available to any interested researcher. Researchers are also encouraged to contribute their own research reports to the catalogue to strengthen the utility of the site.

Table 1. Pretesting methods, their strengths, and weaknesses.

	Pretesting Method	What it is	Strengths	Weaknesses	Most Common Use
Field Methods	Field pilot study (for an overview, see [12])	A miniature version of the main data collection	realistic; allows for testing all field procedures; allows for feedback from interviewers, field managers, respondents, and data analysts	costly; requires large sample size relative to the other techniques, needs to be planned and conducted in advance to allow time for changes	field work test
	Interviewer debriefings (for an overview, see [11])	Small group discussion with interviewers to talk about their experiences	uses interviewers' expertise on what makes a question difficult in a particular situation and with particular types of respondents	interviewers themselves may be responsible for the respondents' confusion/problem with a question	field work test
	Respondent debriefings	Respondents' comments on specific questions or the survey as a whole (usually collected during a field pilot study as a separate interview);	cheap - conducted as part of the field pilot study; allows for identification of question-specific problems; large sample size allows for confidence in results; realistic (field setting)	in some cultures, respondents may not want to admit confusion and inability to understand a question; increases respondent burden as the length of the interview increases; may be hard to recall items that were problematic	field work test
	Behavior coding (e.g., [16]; also, [12])	Systematic coding of the interviewer-respondent interaction in order to identify problems that arise during the question-answer process	direct observation of the question-answer process; comparability when standard codes are employed; replicable; allows for use of universal codes, but also study specific; quantitative; requires medium sample size (30 interviews are considered sufficient to detect problems)	time and labor intensive; requires well trained coders and consistent use of the coding scheme; does not identify the exact problem in a question with many codes	questionnaire testing; field management

	Focus groups (see [6] for an overview; also [12])	Small group of people brought together to discuss specific topics in a relatively unstructured manner, led by a moderator who ensures the flow of the conversation is in the intended direction	useful when there is no information on the topic of interest; uses the same types of respondents who are the target population for the survey; allows for immediate follow up; requires small group size (10-12 participants)	mainly qualitative; results should be carefully interpreted due to small sample size; requires well trained moderators; small group dynamics may influence the results	questionnaire development
Cognitive Laboratory Methods (for an overview, see [11])	Vignettes (e.g., [24])	Brief stories/scenarios describing hypothetical situations or persons and their behaviors to which respondents are asked to react in order to allow the researcher to explore contextual influences on respondent's response formation processes	allows for quantitative analyses; suitable for sensitive topics; requires small sample size relative to the other techniques	disconnect between a hypothetical situation and respondent's actual views and behaviors; cultures may differ in their ability to think hypothetically (e.g., [3])	questionnaire development; concept understanding test
	Concurrent think-aloud [2] [6].	Respondents' report of the thoughts they are having while answering a survey question	open format with potential for unanticipated information; lack of interviewer bias when probes are not used	unnatural; high respondent burden; may affect the natural response formation process, thus provide unrealistic picture of how respondents answer questions in the field; coding may be burdensome; assumes respondents are able to indentify and report what information they used to come up with a response to the survey question; respondents may begin to overinterpret the questions and come up with problems that do not exist in the natural context	questionnaire development

	Retrospective think-aloud [1]	Interview with respondents after they have completed a survey about how they came up with answers to specific questions	does not interfere with the response formation process	assumes respondents are able to indentify and report what information they used to come up with a response to the survey question; assumes information is still available in short-term memory	questionnaire development
Other	Expert review (for an overview, see [12])	Review of draft materials by experienced methodologists, analysts, translators	cost efficient; quick; can identify a wide variety of problems in the survey questionnaire (from typos to skip patterns); requires very small sample of experts (usually 2-3)	subjective; no "real" respondents involved	questionnaire development
	Question Appraisal System (for example, [28])	A systematic appraisal of survey questions that allows the user to identify potential problems in the wording or structure of the questions that may lead to difficulties in question administration, miscommunication, or other failings.	cost efficient; provides sense of reliability due to standardization	identifies a problem without pointing out to a solution	questionnaire development
	Usability Testing [13] [27]	Testing of the functionalities of CAPI, CATI, sample management systems or printed materials such as respondent and interviewer booklet, show cards, etc.	direct user assessment of the tools that will be used during data collection; can be cheap - can be conducted with employees of the survey organization; usually requires small sample sizes	time consuming;	field work test

Statistical Modeling	Multi-trait-multi-method (MTMM) Database [25]	Database of MTMM studies that provides estimates of reliability and validity for over 1000 questionnaire items	provides quantitative measures of question quality	costly and labor intensive; questions are considered in isolation, so question order effects might be ignored	questionnaire development
	Item Response Theory (IRT) Approach [23]	Statistical models that allow to examine how different items discriminate across respondents with the same value on a trait	provides a quantitative measure of item functioning; suitable for scale development	requires data collection; questions considered in isolation	questionnaire development

Glossary

Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Attitudinal question	A question asking about respondents' opinions, judgments, emotions, and perceptions. These cannot be measured by other means; we are dependent on respondents' answers. Example: *Do you think smoking cigarettes is bad for the smoker's health?*
Audio computer-assisted self-interviewing (ACASI)	A mode in which the respondent in which the respondent uses a computer that plays audio recordings of the questions to the respondent, who then enters his/her answers. The computer may or may not display the questions on the screen.
Behavior coding	Systematic coding of the interviewer-respondent interaction in order to identify problems and sometimes to estimate the frequency of behaviors that occur during the question-answer process.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: *Do you smoke?* *Yes* ___ *No* ___
Cognitive interviews	A pretesting method designed to uncover problems in survey items by having respondents think out loud while answering a question or retrospectively.

Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Computer assisted personal interviewing (CAPI)	A face-to-face interviewing mode in which a computer displays the questions onscreen, the interviewer reads them to the respondent, and enters the respondent's answers directly into the computer.
Computer assisted recorded interviewing (CARI)	A system for audio recording of interviews (or interview parts) that allow for monitoring interviewer performance in the filed/call center and detection of data fraud.
Computer assisted telephone interviewing (CATI)	A telephone interviewing mode in which a computer displays the questions on a screen, the interviewer reads them to the respondent over the phone, and enters the respondent's answers directly into the computer.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written consent form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Context effects	The effect of question context, such as the order or layout of questions, on survey responses.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.

Differential item functioning (dif)	Item bias as a result of systematic differences in responses across cultures due to features of the item or measure itself, such as poor translation or ambiguous wording.
Double-barreled (questions)	Survey questions that inadvertently ask about two topics at once.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Focus group	Small group discussions under the guidance of a moderator, often used in qualitative research that can also be used to test survey questionnaires and survey protocols.
Item nonresponse, item missing data	The absence of information on individual data items for a sample element where other data items were successfully obtained.
Item Response Theory (IRT)	A theory that guides statistical techniques used to detect survey or test questions that have item bias or differential response functioning (see dif). IRT is based on the idea that the probability of a response an individual provides is a function of the person's traits and characteristics of the item.
Loaded questions/words	Questions that are worded in such a way that invite respondents to respond in a particular way.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement equivalence	Equivalence of the calibration system used in the questionnaire and the translation.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Mode	Method of data collection.

Multi-Trait-Multi-Method (MTMM)	A technique that uses the correlations between multiple methods (i.e. modes) and multiple traits (i.e. variables) to assess the validity of a measurement process.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Open-ended question	A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers. Example: *What is your occupation?* *Please write in the name or title of your occupation*___________
Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primacy	Context effects in which the placement of the item at the beginning of a list of response options increases the likelihood that it will be selected by the respondent.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.

Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Recency	Context effects in which the placement of the item at the end of a list of response options increases the likelihood that it will be selected by the respondent.
Reliability	The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.
Response distributions	A description of the values and frequencies associated with a particular question.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sample management system	A computerized and/or paper-based system used to assign and monitor sample units and record documentation for sample records (e.g., time and outcome of each contact attempt).
Sample person	A person selected from a sampling frame to participate in a particular survey.

Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Standardized interviewing technique	An interviewing technique in which interviewers are trained to read every question exactly as worded, abstain from interpreting questions or responses, and do not offer much clarification.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Tailoring	The practice of adapting interviewer behavior to the respondent's expressed concerns and other cues, in order to provide feedback to the respondent that addresses his or her perceived reasons for not wanting to participate.
Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample data.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Validity	The extent to which a variable measures what it intends to measure.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.

Vignettes Brief stories/scenarios describing hypothetical situations or persons and their behaviors to which respondents are asked to react in order to allow the researcher to explore contextual influences on respondent's response formation processes.

References

[1] Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot, UK: Gower.

[2] Bickart, B., & Felcher, M. (1996). Expanding and enhancing the use of verbal protocols in survey research. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining the cognitive and communicative processes in survey research* (pp.115-142). San Francisco, CA: Jossey-Bass,.

[3] Bloom, L. (1981). *The linguistic shaping of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[4] Bolton, R., & Bronkhorst, T. (1996). Questionnaire pretesting: Computer-assisted coding of concurrent protocols. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco, CA: Jossey-Bass.

[5] Davis, D. K., Blair, J., Crawley, E. L., Craig, K. M., Rappaport, M. S. B., Baker, C. A., & Hanson, S. L. (2001). *Census 2000 quality survey instrument pilot test* (Report submitted by Development Associates, Inc., Arlington, VA). Washington, DC: US Census Bureau, Center for Survey Methods Research.

[6] Davis, W., & DeMaio, T. (1993). *Comparing the think-aloud interviewing technique with standard interviewing in the redesign of a dietary recall questionnaire*. Alexandria, VA: ASA Section on Survey Research Methods.

[7] Fowler, F. J. Jr. (1995). *Improving survey questions: Design and evaluation* (Vol. 38, Applied social research methods series). Thousand Oaks, CA: Sage Publications.

[8] Gfroerer, J., Eyerman, J., & Chromy, J. (Eds.) (2002). *Redesigning an ongoing national household survey: Methodological issues* (DHHS Publication No. SMA 03-3768). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

[9] Goerman, P. L. (2006). An examination of pretesting methods for multicultural, multilingual surveys: The use of cognitive interviews to test Spanish instruments. In J. A. Harkness (Ed.), *Conducting cross-national and cross-cultural surveys* (ZUMA Nachrichten Spezial Band 12) (pp. 67-79). Mannheim, Germany: ZUMA. Retrieved April 14, 2010, from http://www.csdiworkshop.org/pdf/2005_meeting/znspezial12.pdf

[10] Goerman, P. L. & Caspar, R. (2010). Managing the cognitive pretesting of multilingual wurvey instruments: A case study based on the

pretesting of the U.S. Census Bureau bilingual Spanish/English questionnaire. Chapter 5 in J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B-E. Pennell, T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts*. New York, NY: John Wiley & Sons.

[11] Goerman, P. L,, Caspar, R., Sha, M., McAvinchey, G., & Quiroz, R. (2007). *Census bilingual questionnaire research* (Final round 2 report). Retrieved June 3, 2008, from http://www.census.gov/srd/papers/pdf/ssm2007-27.pdf

[12] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[13] Hansen, S. E., & Couper, M. P. (2004). Usability testing as a means of evaluating computer assisted survey Instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al (eds), *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley & Sons.

[14] Helitzer-Allen, D., Makhambera, M., & Wangel, A-M. (1994). Obtaining sensitive information: The need for more than focus groups. *Reproductive Health Matters, 3*, 75-82.

[15] Hunt, S.D., Sparkman, R. D., & Wilcox, J. B. (1982). The pretest in survey research: Issues and preliminary findings. *Journal of Marketing Research 19*, 269-273.

[16] Mangione, T., Fowler, F. J., & Oksenberg, L. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics 8*(3), 293-307.

[17] Martin. E., & Gerber, E. (2004). *Results of recent methodological research on the Hispanic origin and race questions*. Paper presented to the Population Association of America, Boston, MA.

[18] Miller, K., Fitzgerald, R., Caspar, D., M., Gray, M., Nunes, C., Padilla, J.L., et al. (2008). *Design and analysis of cognitive interviews for cross-national testing*. Paper presented at the International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts, Berlin, Germany.

[19] Pan, Y. (2004). Cognitive interviews in languages other than English: Methodological and research issues. *Proceedings of the section on survey research methods*. Alexandria, VA: American Statistical Association.

[20] Pan, Y., Craig, B, & Scollon, S., (2005). Results from Chinese cognitive interviews on the Census 2000 long form: Language, literacy, and cultural issues. *Statistical Research Division's Research Report Series* (Survey Methodology #2005 - 09). Washington, DC: U.S.

Bureau of the Census.

[21] Pan, Y., Landreth, A., Park, H., Hinsdale-Shouse, M. & Schoua-Glusberg, A. (2010). Cognitive interviewing in non-English languages: A cross-cultural perspective. Chapter 6 in J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B-E. Pennell, T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts.* New York, NY: John Wiley & Sons.

[22] Redline, C., R. Smiley, M. L., DeMaio, T., & Dillman, D. A. (1999). Beyond concurrent interviews: An evaluation of cognitive interviewing techniques for self-administered questionnaires. Paper presented at Proceedings of American Statistical Association: Survey Research Methods Section

[23] Reeve, B., & Mâsse, L. (2004). Item Response Theory (IRT) modeling for questionnaire evaluation. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley & Sons.

[24] Rossi, P., & Anderson, A. (1982). The factorial survey approach: An introduction. In P. Rossi & S. Nock (Eds.), *Measuring social judgments: The factorial survey approach*. Thousand Oaks, CA: Sage Publications.

[25] Saris, W., van der Veld, W., & Gallhofer. I. (2004). Development and improvement of questionnaires using predictions of reliability and validity. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley & Sons.

[26] Snijkers, G. (2002). *Cognitive laboratory experiences: On pre-testing computerized questionnaires and data quality*. Heerlen: Statistics Netherlands.

[27] Tarnai, J., & Moore, D. (2004). Methods for testing and evaluating computer-assisted questionnaires. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley & Sons.

[28] Willis, G. B., & Lessler, J. T. (1999). *Question appraisal system BRFSS-QAS: A guide for systematically evaluating survey question wording* (Report prepared for CDC/NCCDPHP/Division of Adult and Community Health Behavioral Surveillance Branch). Rockville, MD: Research Triangle Institute. Retrieved April 14, 2010, from http://appliedresearch.cancer.gov/areas/cognitive/qas99.pdf

Internet Links

[29] European Social Survey - Round Five. Retrieved July 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=80&Itemid=365

[30] Survey of Health, Ageing and Retirement in Europe. Retrieved April, 2011, from http://www.share-project.org/t3/share/uploads/tx_sharepublications/CH_7.1.pdf

XII. Data Collection

Beth-Ellen Pennell, Rachel Levenstein, and Hyun Jung Lee

Introduction

Collecting comparable data in multiple nations and cultures is a highly complex task, in which one can expect to encounter a variety of languages and cultural contexts. Even in a single country, the target population may not be a linguistically, ethnically or culturally homogenous population but a collection of language and cultural groups. Some of the languages involved may not even have a standard written form. The study may also need to take into account wide variations in respondent literacy. The geographic topography may be difficult (e.g., remote islands, deserts, or mountainous regions). Weather and seasonal impediments (e.g., winter/summer, monsoons) may make the harmonization of fielding times across different countries impractical. Some populations may be inaccessible because of migration patterns or only accessible under special circumstances (e.g., miners in camps, or populations in which part of the population goes on long hunting or fishing trips). Other individuals may have refugee or undocumented status. People living in shanty-type housing may not be included on a given sampling frame. Interviewing in some areas may be dangerous for interviewers. While homeless populations are often not included by definition, the number and definition of the “homeless” may differ considerably from location to location. Outside events such as natural disasters or political upheavals may also pose major challenges for data collection.

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 be inappropriate or not feasible in some situations. In addition, the size and composition of nonresponse will likely vary due to differences in contactability and cooperation. Some countries officially prohibit survey research (e.g., North Korea and Burma/Myanmar) or, to date, severely restrict data collection on some topics, or restrict publication of results (e.g., Iran) [35].

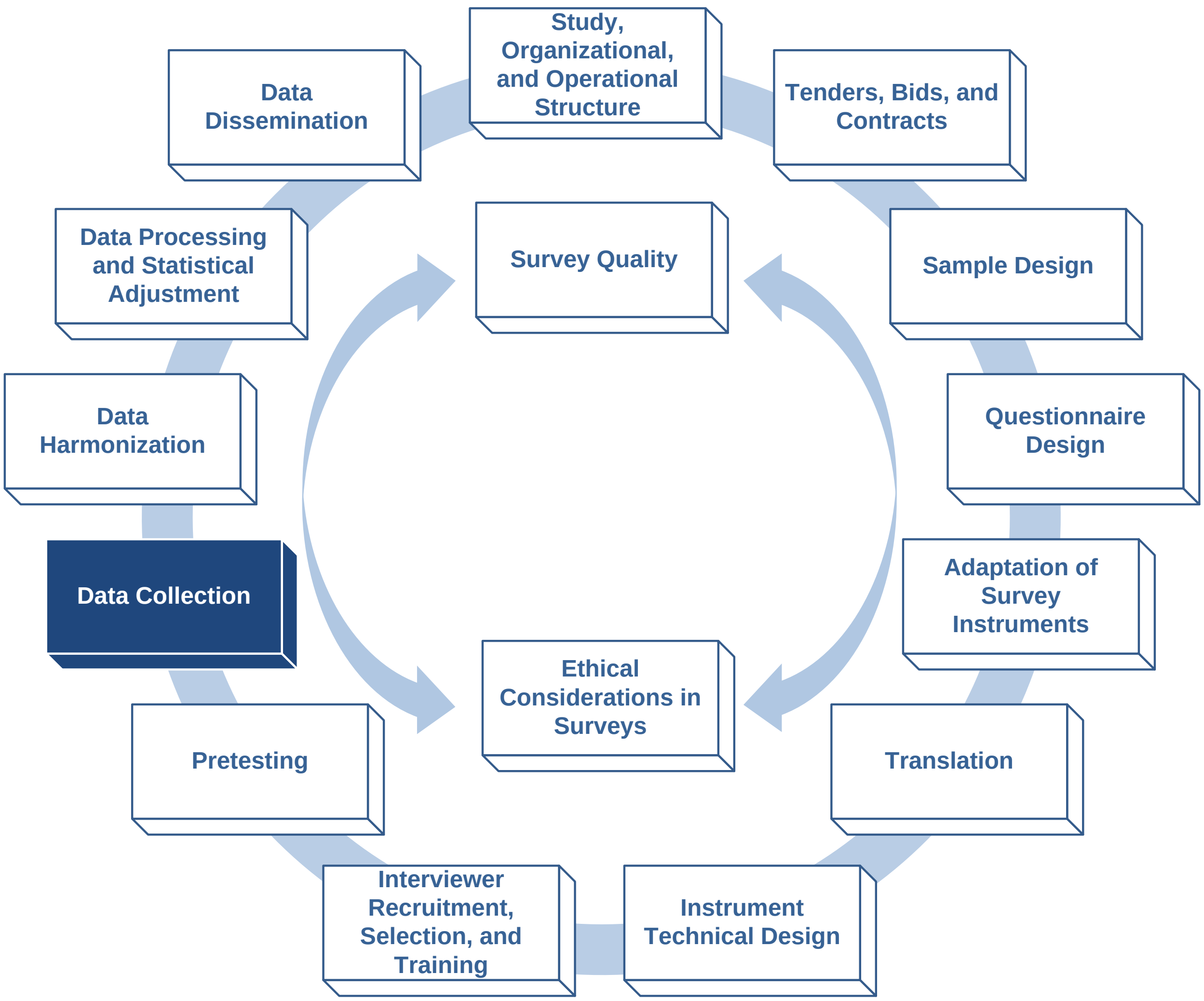
While a survey conducted in a single country might face one or more of the challenges mentioned above, the probability of encountering multiple hurdles is much higher in a large-scale, cross-national study. What is atypical in the one-country context often becomes the norm in cross-national contexts. Moreover, the assumed homogeneity and common ground that may, broadly speaking, hold for a one-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, allowing some flexibility in data collection protocols can reduce costs and error.

These guidelines are intended to advise data collection decision-makers. 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, Organizational, and Operational Structure for more details.

Because difficulties in data collection can be extreme in majority countries, these guidelines heavily emphasize the challenges of data collection in such contexts.

Figure 1 shows data collection within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



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.

1. **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 the population under study can help assess concerns and suggest potential solutions.

Procedural steps

- Assess for each group the appropriateness of (1) the constructs examined, (2) the administration procedures, and (3) the operationalizations [61].

- Plan a pilot study, gathering input from the coordinating center on topics to be included. These might include the survey topic, intended mode of administration, respondent burden (e.g., length of interview, complexity of topic), and proposed methods of dealing with nonresponse.

- Gather information from people who are familiar with data collection in the area. 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.

- Solicit the help of local collaborators or researchers.
  - Local collaborators or researchers can conduct the pilot studies.
    - Provide them with a detailed description of the protocol, including the proposed mode, nonresponse reduction techniques, timing, interviewer training, remuneration, monitoring, and the general framework for data collection.
    - Request feedback on all aspects of the proposed study.
  - Elicit information from these sources and any relevant administrative bodies on:
    - Population issues (e.g., local knowledge about the survey, family structure and household listing issues, literacy levels, and cultural norms).
    - Logistical issues (e.g., seasonal accessibility and dangerous areas).
    - Issues related to mode choice (see Guideline 2).
    - Human protection issues (e.g., legal and cultural permissions which may be necessary to conduct the study) (see Ethical Considerations in Surveys).

Lessons learned

- One-on-one interviews are useful because participants in focus groups may be unwilling to express objections to the data collection protocol in a group setting (see Pretesting).

- 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 the experts believed the populace would think that the interviewers were agents of the government. 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 [13].

2. Select a mode of administration that is appropriate for the survey topic and feasible for the country or culture.

Rationale

Whether dictated by the coordinating center or left to individual survey organizations (see Study, Organizational, and Operational Structure), selecting the mode(s) in which the survey will be administered is a major design decision. It affects comparability, survey cost, survey error, instrument design, and data collection plans. There is no one “best” mode; rather, mode(s) should be chosen based on appropriate tradeoffs of time, cost and error. In an international setting, cultural norms, literacy levels, and logistics will further constrain mode selection.

Surveys can be conducted in numerous ways: face-to-face (FTF), by telephone (either conducted by an interviewer or using Interactive Voice Response (IVR)), through the mail, or over the web. The survey instrument can be paper-and-pencil or computer assisted in format. It can be interviewer-administered or self-administered. This guideline will focus primarily on face-to-face, telephone, and mail modes. Less research has been conducted on IVR, web surveys, or other, newer modes in cross-cultural settings than in single-country surveys. In addition, we have no strong sense of their current viability in multiple contexts around the world. More methodological research is needed in this area.

Procedural steps

- Assess the advantages and disadvantages of each mode. The general considerations for mode choice are [11]:
  - Costs:
    - Generally, mail and web surveys are the least expensive to

implement, followed by phone interviews while face-to-face surveys are the most costly.

- Speed of completion:
  - Telephone surveys can generally be completed in the shortest period of time.
- Complexity of concepts to be measured:
  - Face-to-face interviews and mail surveys can include visual aids.
  - Face-to-face surveys allow for the collection of additional data, such as assessments, physical measures, biomedical samples, and observations of the household, neighborhood, etc.
  - Interviewers in either telephone or face-to-face surveys can help respondents understand complex concepts and probe for complete answers.
- Target population and sampling frame (see Sample Design):
  - Often the sampling frame(s) available will dictate the choice of mode. For example, if an up-to-date sampling frame with names and addresses is available, mail or face-to-face surveying are feasible. If there are no preexisting frames, the survey organization may have to construct an area-based frame.
  - If a large percentage of the population has telephone service, a telephone survey may be feasible.
- Sample dispersion:
  - If the target population spans a wide geographic area, the cost of contacting respondents in person may be prohibitive; in this case, mail or telephone surveys are alternatives.
  - If the sample is widely dispersed and mail or telephone surveying are not viable options, consider clustering to reduce interviewer travel costs (see Sample Design).
- Interest in the topic:
  - If respondents do not seem interested in the topic, telephone or face-to-face interviewers can explain the purpose of the study, answer questions, address concerns, and draw attention toward other desirable aspects of the survey (e.g., the survey's contribution to building knowledge in important areas).
- Nonresponse:
  - Face-to-face surveys generally achieve higher response rates and have less nonresponse bias than telephone surveys or, especially, mail surveys.
- Interviewers and sensitive topics:
  - Research indicates that respondents in interviewer-administered surveys, whether telephone or face-to-face, tend to underreport socially undesirable behaviors (e.g., drug use) and overreport socially desirable behaviors (e.g., voting, religious service attendance).
- Instrument:

    - Complex questionnaires with branching and fills are difficult for respondents to complete on their own. Interviewers may be trained to use questionnaires with branching and fills but may still make errors. Consider using an interviewer-administered mode and a computer-assisted instrument (CAI) with programmed skip patterns. However, using CAI increases the fixed costs and the time needed to develop the survey instrument.

- Use the findings from the focus groups and one-on-one interviews as a guide (see Guideline 1) to determine what particular challenges the specific culture or country presents, including:
  - Access to technology.
    - If the target population has limited access to telephone, then this mode is not appropriate. See Appendix A for telephone (and internet) penetration rates across the world.
    - In some areas, telephone and mail service may be unreliable.
    - Some areas may lack electricity, posing a challenge for computer-assisted interviewing.
    - In some areas the climate might be adverse for computers (heat/cold, humidity).
    - In some areas carrying expensive equipment might put the interviewers in danger of being robbed.
  - Literacy levels of the target population.
    - Written materials are difficult to use in countries or cultures with low literacy levels or in areas where the survey process is unfamiliar (for example, using scales).

- Consider using a combination of modes to solve design problems while achieving an optimal data collection [19].
  - Problems which may be addressed include reducing coverage bias by expanding the sampling frame, decreasing nonresponse bias, and reducing social desirability bias.
  - Modes can be combined in any of several ways [21]:
    - Multiple modes can be used concurrently within a culture or country. For example, households that can be easily reached by telephone could be surveyed on the phone while the remainder is interviewed face- to-face.
    - Different modes can be used concurrently in different cultures or countries. A specific mode may be optimal for one country or culture while a different mode is the best choice for another.
    - Different modes can be used sequentially. The survey can be administered primarily in one mode with additional modes offered as part of a nonresponse follow-up program. Similarly, the first wave of a panel survey could be conducted face- to-

face, with subsequent waves switching to a less expensive mode.

- It is crucial to understand the cost and error implications prior to implementing a mixed mode design.
  - Using a combination of modes can affect data comparability because the sources and level of error differ from one mode to another (mode effects), and these mode effects may differ across country or culture.
  - Planning decisions regarding the standardization of survey specifications may dictate one mode for all populations.
  - Ideally, conduct a pilot study to estimate cross-national mode effects. Using a subsample, randomize the mode and compare key survey estimates to measure the impact of mode on these estimates (see Guideline 10; see [40] [58] for details on cross-national mode study).

Lessons learned

- Mode choice of a survey organization or coordinating center in cross-cultural surveys is driven by a combination of optimal data collection procedures and country- or culture-specific considerations. For example, in some minority countries, telephone coverage is very high making a telephone survey a possible choice. In general, the response rate in a telephone study can be expected to be higher than the response rate using a mail survey. Total cost should be lower than in a face-to-face survey. However, quality and precision might also be lower due to nonresponse bias or if the majority of the target population is not literate or a reliable telephone service is unavailable. In these latter cases, face-to-face interviewing may be the only feasible mode [58].

- Many large established cross-cultural surveys (e.g., the Afrobarometer Survey [63], the Asian Barometer [64], the Living Standard Measurement Study Survey [49], the World Value Survey [69]) have chosen face-to-face interviewing. Round 5 of the European Social Survey [66] and the Survey of Health, Ageing and Retirement in Europe [68] conduct face-to-face interviews along with a supplementary self-administered questionnaire. The International Social Survey Programme [67] survey is self-administered.

- While a mixed-mode design can reduce the cost of data collection, it may also create an additional layer of complexity and, thus, the overall costs for coordinating centers. The Gallup World Poll implements a mixed mode design in which the telephone is used in countries where80% 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 [11], so overall data collection costs are reduced. However, comparability problems due to different modes (phone in one country, face-to-face in another) may exist [26].

- In a cross-national context, the impact of mode can be confounded with cultural differences. For example, the ISSP originally planned to have all countries use a self-administered mode. However, low literacy levels in some countries necessitate the use of interviewers [58]. Differences between countries on survey estimates may be substantive or may be a result of mode effects.

3. If face-to-face interviewing is selected, establish procedures for dealing with issues specific to this mode.

Rationale

Many cross-cultural projects attempt to keep the mode of administration constant by choosing face-to-face data collection, as it generally has the best sample coverage properties, highest response rates (and therefore possibly lower nonresponse bias), and does not require respondents to be literate. In order to collect comparable data, surveys that are conducted in multiple countries or cultures must establish a standard data collection protocol. At the same time, the protocol will sometimes need to allow for modifications required by local norms, conditions or customs.

Telephone, mail, and even web modes may be used in some cross-cultural surveys. However, the implementation of face-to-face surveys presents a number of logistical challenges not faced in other modes.

Procedural steps

- Contact local authorities for clearance to collect data at the sampled site; if necessary, negotiate with local authorities or militias to gain access to sample areas.

- Take measures to ensure interviewer safety.
  - Inquire about potential safety problems, such as civil unrest and high crime areas.
  - Decide whether interviewers should travel in groups and be accompanied by security personnel.
  - If it is safe, have interviewers visit their work areas during the day on or before the first day of data collection. They should check for potential hazards and safe havens during this visit.

  - Have interviewers tell their supervisors and family members when they plan to leave for the field, the location of the area, and when to expect them back.

- Have interviewers carry the following items in the field to establish their legitimacy:
  - Official identification from the survey organization.
  - Official letters to local authorities describing the study.

- Provide adequate transportation for staff and supplies.
  - If maps are unavailable or unreliable, consider the use of local guides or GPS instruments.
  - Arrange to secure fuel and oil and to maintain the vehicles used by the field staff; this may present logistical problems in some majority countries.
  - Arrange for emergency transportation in the event that a field team member becomes ill or injured and needs immediate medical attention or it becomes unsafe to stay in an area.
  - Arrange for backup transportation.
  - Secure housing accommodations in more remote areas prior to fieldwork or have the team carry their own (e.g., tents or mobile homes).

- Match interviewer and respondent characteristics (e.g., race, ethnicity, or gender) when cultural norms so dictate (see Interviewer Recruitment, Selection, and Training).

- If physical measurements are taken as part of the survey,
  - Check the cultural acceptance of taking such measurements.
  - Calibrate the equipment regularly.

- Provide all members of the field staff with access to a reliable line of communication with their supervisor. This will allow them to report progress and problems and to transmit the survey data as quickly as possible.
  - Majority countries may have weak communication capacities, especially in rural areas.
  - Cellular or satellite phones may be a worthwhile investment for teams in remote areas.

- Aim to conduct the interview in a setting which affords visual, physical, and auditory privacy.
  - Privacy is critical for keeping respondents' answers to the survey questions confidential.
  - Although complete privacy is ideal, it is impossible to achieve in

some cultures. Interviewers should attempt to keep the interview as private as possible, while still respecting cultural norms. This may involve self-administration on more sensitive questions. Another alternative may be to keep any others present occupied while the targeted respondent completes the survey.

- Customs may vary among countries. In some, it may be unacceptable to have any interviewer come to the respondent's home, or it may be unacceptable for an interviewer of opposite sex to the selected respondent or informant to enter the home.
- Privacy increases the likelihood that respondents will answer honestly about sensitive behaviors, such as sexual practices or drug use. What is considered sensitive may vary among countries or cultures; administration practices may need to differ accordingly.

Lessons learned

- Because responses to some survey questions can be affected by other individuals present during data collection, it is optimal—but not always possible—to conduct face-to-face surveys in private. In a face-to-face fertility survey of women in what is now Bangladesh, privacy was difficult to establish; most interviews took place in the presence of the respondent's mother- or sister-in-law. This may have affected responses to sensitive questions [15].

- Similarly, men in some parts of Africa were found to object to confidential interviews with their wives or children. The interviewers were instructed to conduct interviews in a place that was visible to the male heads of household but out of earshot [14].

- In some rural places it might not always be feasible to conduct a survey inside a home. During the Afrobarometer survey, the interview took place outside “under a tree.”

- In some cultures affluent respondents would not allow interviewers in their homes, thus interviews might take place in a café or office.

4. 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. First, applying a combination of research methodologies to study the same phenomenon facilitates the validation of data through cross verification, while each method counterbalances the potential limitations of the others. Second, positive interactions between the interviewer and the respondent in a more unstructured setting may increase the accuracy of the information the respondent provides as well as his or her willingness to provide such information. Finally, qualitative methods can place the behavior of respondents into a broader context and can improve coding by revealing unanticipated influences.

Procedural steps

- Choose data collection methods to fit the aim of the research question [5].

- Consider combining less structured interviewing, field notes, observation, historical materials, or life history calendars with the standardized survey [5].
  - 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 methodological triangulation and integrating qualitative and quantitative methods in data collection, see Further Reading).
  - Triangulation can also widen and deepen one's understanding of the phenomenon being studied.

- 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 [5] [24].

- Remember that traditional qualitative methods can be more expensive and time consuming than a standardized survey [50] [51].

Lessons learned

Several cross-cultural projects have successfully combined qualitative and quantitative methods of data collection.

- 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 [4] [25].

- The life history calendar method is easily adaptable to fit cultural needs. Some tribes in the Chitwin 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 [5] [6] [7].

- Broom [12] 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 holistic understanding of the patients' decision-making processes.

- The Demographic and Health Surveys (DHS) program conducts research in approximately 75 developing countries across the world. The main objectives of the DHS program are "(1) to provide decision makers in the survey countries with data and analyses useful for informed policy choices, (2) to expand the international population and health database, (3) to advance survey methodology, and (4) to develop in participating countries the skills and resources necessary to conduct demographic and health surveys." Phase II of the DHS introduced a calendar at the end of one of the core questionnaires to clarify dates relating to fertility, contraceptive, postpartum, marriage, migration, and employment history. The researchers found that the calendar provided gains in the quantity and quality of data collected, as well as increasing their analytical potential [65].

- Hargreaves et al [33] 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 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 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 [36].

- Keller [41] studied the influence of parents and other socialization factors on human development. Working with young infants 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 footage.

- Implementing qualitative methods helped University of Chicago researcher Douglas Massy gain greater insight into the reasons behind migration in the U.S. [50].

- 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 [54].

5. **Establish a clear protocol for managing the survey sample.**

Rationale

Nonresponse can be assessed and reduced with an effective sample management monitoring system. In addition, a good sample management system facilitates evaluating interviewer workload and performance.

The study structure may specify what sample management systems are used. In cross-cultural surveys with strong centralized control, a single sample management system may be specified in the contract with local survey organizations. If an electronic system is used, coordinating centers may play a role in monitoring fieldwork. See Study, Organizational, and Operational Structure for details.

Procedural steps

- Use a coversheet or an electronic sample management system to track each sample element during the study (see Appendix C for an example of a paper coversheet).
    - Interviewers using paper coversheets have found that they work most efficiently if they sort the coversheets by (1) appointment time and (2) geographical location. The same sorting procedures should be available in electronic sample management systems.
    - Consider efficient methods that allow interviewers to fill in coversheets and do household contacting at the same time. Filling in coversheet forms after making the contact has shown to be error prone.

- Structure the field staff to aid them in working the sample efficiently.
    - Give supervisors the responsibility of assigning sample elements to interviewers and reassigning them when necessary.
    - Do not allow interviewers to switch sample among themselves without the explicit approval of the supervisor.
    - In a face-to-face study, ensure that sample elements are assigned in a way that minimizes travel efforts and costs.
    - In a face-to-face study, decide whether interviewers will work alone, in pairs, or in traveling teams (see above and Interviewer

Recruitment, Selection, and Training).
  - Decide whether interviewers and respondents should be matched on some characteristic(s) (see Guideline 1).
    - If the respondents' characteristics are unknown 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.

- Train interviewers to complete household enumeration and randomly select eligible members within the household unit (see Appendix D for household enumeration).
  - When using computer-assisted interviewing, have interviewers use the computer to randomly select an eligible respondent.
  - When respondent selection cannot be done electronically, have interviewers use a random selection technique, such as a Kish selection table (see Appendix E for an example of a Kish table) [44].

- Have interviewers complete a contact attempt record each time they attempt contact, whether or not the attempt is successful (see Appendix F for an example of a contact attempt record).
  - Use disposition codes to describe the outcome of each contact attempt.
  - Distinguish (1) completed interviews with eligible persons, (2) non-interviews (eligible persons), (3) non-interviews (unknown if eligible persons), and (4) non-interviews (ineligible persons).

- Assign a final disposition code to each sample element in the gross sample at the end of data collection; include any new sample elements that may be created or generated during data collection (e.g., for additional family members or through half open intervals).
  - Provide a clear explanation and training to interviewers before they are allowed to assign final disposition codes.
  - Take into account that, in some survey organizations, only supervisors can assign final disposition codes.

Lessons learned

- An effective sample management system can clarify the causes of nonresponse. When the Amenities and Services 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 study 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 [59].

- Not all survey organizations will be familiar with sample management practices. Allow some time in training for interviewers to become familiar with the sample management system (see Interviewer Recruitment, Selection, and Training) and check completed forms.

6. **Reduce nonresponse bias as much as possible.**

Rationale

Optimal data collections minimize nonresponse bias and maximize response. Increasing the response rate can improve the accuracy of survey data. Although the response rate alone does not predict nonresponse bias [28], 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 and, increasingly, reluctance to participate [20].

Procedural steps

- Minimize nonresponse bias as much as possible.
  - Nonresponse bias is a function of both the response rate and the difference between respondents and nonrespondents on a particular statistic [30]. Because nonresponse bias is statistic-specific, response rates alone do not indicate nonresponse bias.
  - Estimate the effect of nonresponse bias on key survey estimates, if possible (see Guideline 10).
  - If possible, use weighting and imputation [31] (see Data Processing and Statistical Adjustment).

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

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

- Publicize the survey locally to raise awareness and encourage cooperation.
  - If most of the population is literate, consider posting colorful, attractive leaflets.
  - Also use word-of-mouth channels or local dignitaries (doctors, teachers) as appropriate.

- Send prenotification letters to sampled households if feasible.
  - 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, and (5) provide contact information for the organization (see Appendix B for an example of prenotification letters).
  - 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 in Surveys).

- Attempt to contact the respondent.
  - Be aware that different modes face different obstacles.
    - Telephone surveys may encounter caller ID and answering machines; these devices may be used by potential respondents as screening mechanisms, thus preventing contact.
    - Face-to-face surveys must contend with general at-home patterns of sample persons, locked apartment buildings, gated communities, or long absences from home.
    - In mail surveys, it is difficult to disentangle the effects of noncontact, refusal, and a poor sampling frame. Also in a multi-person household it is often not clear who the actual respondent was. If a mail questionnaire is not returned, the respondent may not have picked up mail at that address, may have decided against participating, or may never have received the questionnaire because the address was incorrect.
  - In a face-to-face survey, train the interviewers to make observations of the housing unit to assess likely at-home patterns.
    - In some countries interviewers are not allowed to ask neighbors about targeted but not contacted respondents.
  - In face to face surveys, train the interviewers to use a predefined grid showing different blocks of time across the week when the interviewer must attempt to contact respondents.
    - This practice increases 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 weekends [30].
        - Alternatively, specify the minimum number of times that attempts must be made during daytime hours, during evening hours, and during the weekend (see [46] for details on call scheduling). Incorporate culture-specific information about likely at-home patterns, such as normal workdays, normal work hours, and holidays. Beware of religious and other cultural norms that restrict interviewing at certain times.
    - Specify both the minimum as well as the maximum number of attempts to contact before the final disposition code is assigned to increase efficiency.

- 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.

- Use responsive designs to remedy low contact rates and high nonresponse in specific areas or social strata (see Sample Design) .

- If appropriate, offer an incentive for participation [56].
    - Adapt type and amount of incentive to local custom. Make yourself familiar with country-specific research on incentives.
    - 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 response.
        - Make the token reasonable; it should not be so large that it might raise suspicion about the researcher's or organization's motives or be somehow coercive. It should be generally proportionate to the respondent burden.
        - Ideally, provide the incentive prior to the interview. Incentives promised upon the completion of the interview also increase participation, but to a lesser degree [8] [57].
    - Document the use of incentives, including amount and type, time of implementation, and any special strategy, such as increasing the amount of the incentive in the final weeks of the study.
    - For financial incentives, interviewers may be asked to record that an incentive was given to a respondent; similarly, the respondent may need to sign to indicate receipt.

- Train interviewers to use culturally appropriate reluctance aversion techniques (see Interviewer Recruitment, Selection, and Training).
    - Social or psychological factors (e.g., reciprocation, consistency, social validation, authority, scarcity, liking) affects respondents' decision in survey participation [16]. Minimally, train interviewers how to answer anticipated respondent concerns [29].

  - Be aware that local customs and legal limitations may prohibit any attempt to recontact someone who has declined to participate in the survey. In these cases using appropriate reluctance aversion techniques becomes especially important.
  - Make sure that supervisors monitor interviewers closely on respondent reluctance issues.

- Consider assigning supervisors or more experienced interviewers to cases where interviewers have been unsuccessful making contact or achieving cooperation.

- Consider switching modes to increase contact and cooperation (see Guideline 2).
  - 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.
  - Different modes may produce different survey estimates. These mode-specific differences in measurement might be acceptable to the investigator if nonresponse is sufficiently reduced.
  - If more than one mode is expected to be used and budget permits, examine possible mode effects prior to the 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 end of data collection.
    - Ascertain whether respondents surveyed in each mode produce similar response distributions on key variables before combining their responses for analysis.

Lessons learned

- While the literature has clearly established the positive effects of prepaid and cash incentives upon response in minority countries [8] [57], it is possible that incentives may affect the propensity to respond differently in majority countries. 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.

- Response rates are not necessarily good indicators of nonresponse bias, but nevertheless tend to be used as a proxy 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 [17].

- The same incentive may affect response rates differently across countries or cultures. In the German General Social Survey (ALLBUS), the same incentive (€10) was offered to all respondents. The authors examined cooperation rates for Moroccan and Turkish immigrants. The authors found that the incentive affected cooperation differently by ethnicity and gender: cooperation rates increased as a result of the incentive for Moroccan women, but did not increase for Moroccan men, Turkish men, or Turkish women [62].

Be careful when choosing to give monetary awards to study participants. Keller studied the influence of parents and other socialization factors on human development in Asia, Latin America, Europe, North America, and Africa. Respondents received a cash incentive. Keller experienced some hostility from families that were not selected for the study (and, thus, not given any monetary rewards) because they did have young children [41].

- Countries have different incentive norms. 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]. Some cross-cultural surveys (e.g., the European Social Survey [66], the Living Standard Measurement Study Survey [49]) allow each participating country to decide whether or not to offer incentives. If incentives are offered, they may vary from one country to another. For example, the Survey of Health, Ageing and Retirement in Europe [68] offers various incentives, depending on the country's culture. Incentives for the World Mental Health Survey [43] 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.

- Similarly, many cross-cultural surveys (e.g., the European Social Survey [66], the Living Standard Measurement Study Survey [49], and the World Mental Health Survey [43]) 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 [68], advance letters are mailed to each household in the gross sample and follow-up letters are used with reluctant respondents.

7. **Time data collection activities appropriately.**

Rationale

A specific survey estimate of interest may determine the timing of data collection activities; for example, a survey 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 earthquake 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

- Based upon feasibility studies (see Guideline 1), evaluate environmental, political, and cultural considerations which might affect the timing of data collection. These could include:
  - Extreme weather patterns or natural disasters.
  - War, civil war, military rule, militia rule, or the possibility of hostage taking.
  - Religious and secular holidays or migratory patterns of nomadic people. For example, Independence days (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.

- Establish a specific start and end date for data collection.
  - Keep a concurrent fielding period across countries. This would guarantee the 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 [45].
  - Because unexpected events can interfere with data collection activities, remain somewhat flexible to allow for unexpected events.

Lessons learned

- Coordination of data collection activities across countries or cultures can be difficult or 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 campaigns, rainy seasons, times of famine, and national or religious holidays. Since such events vary across countries and cultures, fieldwork activities are spread over a full year [63].

- 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 must time data collection activities according to local election cycles [37].

- 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 voters' lists or resident registries. As a result, many people mistakenly believed that the new regulation also applied to academic research [38].

- Data collection in Germany for the first European Social Survey had to be delayed due to general elections held in that autumn.

8. 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. Improvement made during data collection may introduce some measure of inconsistency in the data, however. This trade-off should be considered before any action is taken [32].

Procedural steps

- Evaluate the effectiveness of data collection protocols regularly. Include:
  - Sample management systems.
  - Contact protocols.
  - Reluctance aversion protocols.

- Observe the interviewers throughout data collection [47]; monitor them more frequently early in the study, less frequently as the study continues.

- Review a random sample of coversheets on an ongoing basis to ensure that correct eligibility and respondent selection procedures are being followed.

- If the survey is being conducted in a centralized telephone facility, follow established monitoring procedures [18]:
  - Monitor in relatively short (e.g., one-hour) shifts; this is cost-effective and reduces supervisor fatigue.
  - Use probability sampling to ensure that the number of interviews monitored is proportional to the number of interviewers working each hour (see Sample Design).
  - Monitor new interviewers at a higher rate than experienced interviewers.
  - Select from eligible cases in which the phone is still ringing so that the supervisor is not forced to wait for new interviews to begin in order to start monitoring.

- If feasible, audio record face-to-face interviews for review.
  - Determine whether cultural norms permit taping.
  - Be sure to inform respondents that they may be recorded for quality control purposes and allow respondents to refuse to be recorded.
  - Store any tapes safely and securely (see Ethical Considerations in Surveys).

- If physical measurements are being taken, periodically retest the interviewers on the use of these instruments.
  - Select equipment that can withstand the local conditions (heat, cold, altitude, etc).
  - Document the technical specifications of the equipment chosen.

- Re-calibrate equipment as needed throughout data collection Provide interviewers with feedback, both individually and as a group [18] [47].
  - Provide immediate, individual feedback if there has been a critical error.
  - Provide routine, individual feedback for self-improvement.
  - Offer group feedback to focus efforts on improving the process.
  - Evaluate the following with respect to interviewers [11]:
    - 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 (e.g., reading questions as worded, probing, and clarifying).
- 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 and administering tests of physical performance or cognitive ability).

- Whenever possible, recontact or reinterview approximately 10-15% of each interviewer's completed cases, selected at random [1] [53].
  - If recontacting the respondent, verify that the interview took place, inquire if interviewer acted professionally, and ask factual questions (e.g., mode of data collection, interview length, incentive, household composition, and key survey topics) [1].
  - 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 [23] [60].
  - Conduct reinterviews within a time period that is not so long that respondents will have forgotten about the survey or so short that respondents will remember all the details of the survey [23].
  - Make sure recontacts and reinterviews are made with the original respondent and that questions refer to the same time period as that asked about in the original interview [23].

- 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.

- Monitor quality indicators consistently throughout the field period; use an electronic system or note them in a daily log book [60]. Include the following:
  - Distributions of key variables.
  - Hours per interview per interviewer, for the study as a whole, and by respondent groups of interest.
  - Number of respondents approached, interviews completed, incomplete interviews, and contact attempts.
  - Response, refusal, and noncontact rates [60] (see Data Processing and Statistical Adjustment).

  - Outcomes of all contacts and review of disposition code assignment.

- Create statistical process control charts (SPCs) to provide timely information on key aspects of the data collection process [52].
  - Use the charts to detect observations that are not within pre-determined limits (often between one and three standard deviations of the mean).
    - A common use of SPCs in survey organizations is to assess nonresponse reduction methods over the field period. Using these charts, the impact of interviewer effort on response rates can be easily assessed (see case studies in Survey Quality for examples of SPCs).
  - Give extreme observations additional attention and try to determine the root cause.
  - Refer to the charts when deciding whether to release additional sample elements for interviewers to attempt to contact, further monitor interviewers, and offer additional training sessions.

- Set contact limitations, determining:
  - The point at which additional attempts to contact a sample element are inefficient.
  - Whether respondents cooperating after a certain number of contact attempts are significantly different from others on key indicators.

- Identify potential interviewer falsification.
  - Implement silent monitoring in centralized facilities, use audio-recordings and recontacts in field studies, and analyze outliers in the data to detect falsification [1].
  - Check responses to stem questions for each interviewer. Questions that have a stem-branch structure—in which specific responses to “stem” questions require the interviewer to ask a number of “branch” questions—can be 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.
  - Examine paradata, such as keystroke data and time stamps, by interviewer to identify potential falsification.
  - If falsification of data is suspected, contact the respondents involved over the telephone [23]. If respondents cannot be reached via telephone, send out a brief mail questionnaire with a prepaid return envelope [11].
  - If falsification of data is suspected, investigate the interviewer’s other work and remove the interviewer from all data collection activities until the issues have been resolved [1].

- If irregularities or falsified data are discovered, redo the interviewer's cases and delete all of his or her recorded data [1] [11].

Lessons learned

- Process and progress indicators are often interdependent. Therefore, improving one process or progress indicator may negatively affect another. For example, the pursuit of higher response rates can actually increase nonresponse bias if the techniques used to obtain the higher response rates are more acceptable and effective in some cultures than in others [28] [34].

- In Round 4 of the Afrobarometer Survey, teams of four interviewers travel together to the field under the leadership of a field supervisor who has at least an undergraduate degree and experience in collecting data and managing field work teams or no degree but extensive experience. It is the supervisor's job to ensure quality control of survey returns on a daily basis. Interviewers are monitored at all stages and debriefed daily immediately after interviews. Completed questionnaires are checked for missing data and inconsistencies. Each field supervisor maintains a daily written log of observations on sampling and interviewing conditions and political and economic features of the area and and makes daily telephone report to headquarters. A fieldwork debriefing is held after all returns have been submitted. Sampling back-checks are routinely conducted to ensure that the respondent selection is correctly done. The field supervisor also verifies basic information (e.g., respondent age and level of formal education) [63].

- The Asian Barometer survey required all interview teams to travel together under the supervision of a field supervisor, and interviewers to have a debriefing meeting each evening. Supervisors randomly checked with respondents to make sure the interviews were done properly [2].

- In Round 5 of the European Social Survey (ESS), quality control back-checks are performed for at least 10% of respondents and 5% of the non-respondents either in person, by telephone, or by mail. For the respondents, a short interview is conducted to confirm the interview, whether showcards were used, the approximate length of the interview, etc. [66]

- In the Living Standard Measurement Study (LSMS), each field supervisor oversees two interviewers. Each week the field supervisor observes and evaluates one interview per interviewer and documents the process for submission to the national office. Data collection is broken into two rounds; the first half of the questionnaire is completed in round one and then checked for accuracy before the second half of 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 [49].

- The Survey of Health, Ageing 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. The CMS monitors the survey progress in real-time, including screening for eligible respondents, recording contact attempts, enforcing contact and follow-up strategies, and managing refusal conversion strategies. Bi-weekly reports are generated for the SHARE coordinating team [68].

- 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 or all of the interview for 5% to 10% of each interviewer's work. Supervisors randomly select 10% of interviewed households, confirm the household listing and selection procedure, and repeat some of 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 [42][43].

9. 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

- Document the following (see Appendix G):
  - A summary of feedback from the feasibility studies.
  - The interview or data collection process.

- A description of the mode(s) used.
- A description of the mode-specific protocols.
- A description of the sample management system.
- A description of any paradata collected.
- Special approaches to reduce nonresponse, including any incentives and nonresponse follow up.
- Outcome rates by key respondent groups, including response, refusal, noncontact, and other nonresponse rates.
- Structure of the field staff (e.g., size of interviewer groups and supervisor/interviewer ratio).
- Timing of the fieldwork for each country or cultural group.
- 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.
- Any validation study descriptions and outcomes (see Guideline 10)

10. When possible, conduct validation studies to estimate bias.

Rationale

As noted in Guideline 6, response rates alone are not good indicators of nonresponse bias; understanding nonresponse bias and making subsequent post-survey adjustments require information about the nonrespondents. Similarly, measurement error bias can only be estimated when "true" values for survey variables are known or can be modeled (i.e., using latent class analysis). Validation studies can increase confidence in results, assist with post-survey adjustments (see Data Processing and Statistical Adjustment), and address potential criticisms of the study. However, while the interpretation 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. While 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

- Collect data on nonrespondents, if possible, to estimate nonresponse bias [28].
  - 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 design 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 [48].
- Document the data collection procedures, including the proportion of initial nonrespondents included in the validation study, mode of administration, and any additional incentive [32].

- 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.
- 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.
- 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.
- 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 the nonresponse mechanisms at work. While some amount of nonresponse bias may be controlled using these 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.

- Use methodological studies to assess measurement error.
  - 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 in the United States do not use probability-based samples; therefore, errors detected in the self-

selected laboratory sample may not be representative of errors in the target population.

- Another approach is to check outside records for the true value, or a proxy of the true value, of the measure.
  - 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.
- 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 – see discussion in Guideline 6).

- Consider using other methods of assessing measurement error.
  - Reinterview respondents. Reinterviews are especially useful in determining interviewer falsification [23], but may also help assess other forms of measurement error (see [9] [10] for details on estimating simple response variance or bias).
  - 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 designated 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).
  - Collect paradata that may be correlated with measurement error (e.g., number of keystrokes, length of interview).
  - Use interpenetration to estimate correlated response variance due to interviewers.

Lessons learned

- Supplemental studies can be difficult and expensive to implement, but they are useful for validating survey results. For example, a study of discharged patients at a French hospital found no difference in patient satisfaction ratings 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 [27].

- Try to use resources to gain knowledge on bias in an efficient way. Validation studies are expensive but come late. Therefore one should first strive for more preventive measures that hopefully make processes almost error-free. Then paradata should be collected and analyzed so that processes can improve and display a decreased variability. Finally some small-scale validation studies, rather than large ones, should be conducted, and used as input to more long-term 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 validation studies.

Appendix A

Telephone and internet penetration by regions [39]

Regions	Number of fixed line telephones/100 inhabitants		Number of mobile telephones/100 inhabitants		Internet users/ 100 inhabitants	
	2003	2008	2003	2008	2003	2008
Africa	1.4	1.5	5.3	32.5	1.3	4.2
Americas	33.9	30.6	29.1	81.8	30.7	43.7
Asia & Pacific*	11.5	15.7	11.7	36.6	6.0	14.6
Europe	46.4	42.5	73.5	117.9	36.7	57.8

*Data of Asia and the Pacific is for the year of 2002 and 2007 instead of 2003 and 2008

Appendix B

Example of pre-notification letter from ESS 2008 [22]

Logo and Contact information of research oraanization

Logo of research organisation

Autumn 2008

[Dear …..]

European Social Survey 2008

Purpose of the survey

[You have/your address/household has] been selected to take part in an international study on what people think about various important issues affecting [country]. The study is being carried out simultaneously in 25 countries across Europe and will help to find out how much or how little people in different countries share the same views and beliefs. It is being paid for from both [country] and European sources.

Confidentiality of answers

The questionnaire covers a wide range of topics and no special knowledge is needed to answer any questions. Your [name/household/address] has been selected from [sampling frame] by scientific methods to ensure that we get a representative picture of people in [country]. We cannot therefore substitute any [name/household/address]. All information you provide will be treated in strict confidence and will never be linked to your name or address.

Voluntary participation

Most people taking part in the study find it an interesting and enjoyable experience, and we hope that you will too. An interviewer will [visit] you shortly to explain more about the study and, if you agree, will arrange a suitable time for the actual interview, which will be carried out in person. Interviews normally take just under an hour. We certainly hope we can rely on your co-operation.

Meanwhile, if you wish to have any further information about the study, please feel free to contact me on the number above.

Thank you in advance for your help,

xxxx xxxxxx

Information on survey interviewer

The interviewer who will be contacting you is:…………………………………………

Appendix C

Coversheet

SPACE CAN BE USED FOR ADMINISTRATIVE INFORMATION

THIS SPACE USED TO AFFIX LABEL CONTAINING SAMPLE INFORMATION	THIS SPACE USED TO AFFIX LABEL CONTAINING INTERVIEWER INFORMATION

THIS SPACE RESERVED TO RECORD ADDITIONAL INFORMATION ABOUT THE LOCATION OF THE SAMPLE CASE IF NECESSARY (e.g., landmarks or housing unit description if address is unavailable, etc.).	Final Result Code: ____________ Date of Final Result (dd/mm/yyyy): __ / __ / _____ Length of interview: ____________ Total calls: ______ Length of edit:

Appendix D

<u>Household Enumeration Table</u>

	HOUSEHOLD ENUMERATION					RESPONDENT SELECTION		
	11 a. Household Member's First Name	11 b. HH Member's Relationship to Informant	11 c. Sex	11 d. Age	11 e. Language Spoken	11 f. Eligible	11 g. Person Number	11 h. Selected R
MALES			M					
			M					
			M					
			M					
			M					
			M					
			M					
FEMALES			F					
			F					
			F					
			F					
			F					
			F					
			F					

Instructions for Household Enumeration Table

Column 11a (Household Member's First Name): List all members of the household, beginning with the informant. Note that males are listed in the upper portion of the table and females in the lower portion.

Column 11b (Household Member's Relationship to Informant): Record each household member's relationship to the informant (e.g., husband or wife, son or daughter, mother or father, brother or sister, friend, etc.).

Column 11d (Age): Record each household member's age.

Column 11e (Language Spoken): This column may or may not be included, depending upon the study requirements.

Column 11f (Eligible): Place a check mark in this column if, based upon the information in columns 11a-11e, the household member meets the eligibility criteria for the study.

Column 11g (Person Number): Assign a sequential number to each eligible household member. Begin by numbering eligible males from oldest to youngest, continue by numbering eligible females from oldest to youngest.

Column 11h (Selected R): Count the number of eligible persons in the household. Find that number in the Kish table in the "If the Number of Eligible Persons is:" column. The selected respondent will be the household member with the "Person Number" corresponding to the "Interview the Person Numbered:" column in the Kish table.

Appendix E

Random Within House Selection Techniques

Kish (1949) selection tables [44]

SELECTION TABLE A

If the number of eligible persons is:	Interview the person numbered:
1	1
2	1
3	1
4	1
5	1
6 or more	1

SELECTION TABLE B_1

If the number of eligible persons is:	Interview the person numbered:
1	1
2	1
3	1
4	1
5	2
6 or more	2

SELECTION TABLE B_2

If the number of eligible persons is:	Interview the person numbered:
1	1
2	1
3	1
4	2
5	2
6 or more	2

SELECTION TABLE C

If the number of eligible persons is:	Interview the person numbered:
1	1
2	1
3	2
4	2
5	3
6 or more	3

SELECTION TABLE D

If the number of eligible persons is:	Interview the person numbered:
1	1
2	2
3	2
4	3
5	4
6 or more	4

SELECTION TABLE E_1

If the number of eligible persons is:	Interview the person numbered:
1	1
2	2
3	3
4	3
5	3
6 or more	5

SELECTION TABLE E_2

If the number of eligible persons is:	Interview the person numbered:
1	1
2	2
3	3
4	4
5	5
6 or more	5

SELECTION TABLE F

If the number of eligible persons is:	Interview the person numbered:
1	1
2	2
3	3
4	4
5	5
6 or more	6

The tables provide unbiased estimates by giving each respondent a weight based on the number of adults in the household. This guarantees that the selection within a household is random for a combined total random sample across the housing units (addresses) that were randomly selected in the first place.

Instructions for Kish tables

1. Assigning Kish tables to the sample file: one of the twelve tables is randomly selected and assigned to the first line in the sample file. The series of twelve is then run through twice, assigning tables to the sample lines. Then again a table is randomly selected and the series is run through twice. This procedure is repeated until all sample lines have an assigned Kish table.

2. Household listing: a household listing of eligible adults (age 18 and over) who reside in that household is taken at each of the sample addresses. Usually the males are listed first in order of decreasing age, and then the females in the same order.

3. Using Kish tables: the table assigns a number to each member of the household listing. Sample Kish tables are shown above. In the first column the interviewer would circle the total number of eligible persons. The corresponding number in the second column of the Selection Table denotes the person selected to be interviewed.

Appendix F

Contact attempt record

	CALL #1	CALL #2	CALL #3	CALL #4
DATE:				
DAY OF WEEK:				
EXACT TIME BEGAN:				
IWER ID:				
CONTACT WITH:	R / INF/ NO ONE	R / INF/ NO ONE	R / INF/ NO ONE	R / INF/ NO ONE
MODE OF CONTACT:	PERSONAL / TEL	PERSONAL / TEL	PERSONAL / TEL	PERSONAL / TEL
TELEPHONE NUMBER IF OBTAINED:				
HU LISTING OBTAINED:	YES / NO	YES / NO	YES / NO	YES / NO
DETAILED DESCRIPTION OF CONTACT OR CONTACT ATTEMPT				
DISPOSITION CODE:				

R = Respondent HU = Housing Unit

Inf = Informant Listing = enumeration

Appendix G

Documentation

	Details	Examples from SMDS[1]	Examples from ISSP[2]
Data collection organizations	• The number of organizations • Contact information • Type of organizations (e.g., government agency, private research company)	How many organizations conducted data collection for this study in your country? If your agency/organization contracted with another organization which provided data collection services, please include that here. ____ organizations	Please enter the name of your institute and your country: Institute: ______________ Country: ______________

[1] Survey Metadata Documentation System (SMDS): a standardized web-based documentation tool which was developed by the University of Michigan's Institute for Social Research and Gesis-ZUMA.

[2] International Social Survey Programme (ISSP): see [55] for details.

Data collection methods	• The number of separate survey data collection efforts • A brief title of each survey data collection efforts • Delivery of sample to interviewers (e.g., computerized sample management system) • Mode of data collection • Screening/respondent selection procedure	How were the face-to-face interviews administered in this study? Please check all that apply. - Computer-assisted personal interviewing (CAPI) - Paper and pencil interviewing (PAPI) - Other, specify: ____________________	What selection method was used to identify a respondent? Please tick all that apply. (do not answer if your sampling frame consists of named individuals – which are the target persons. Then continue with question 66) Kish grid ____ Last (or next) birthday _____ Quota _____ Other (please write in details) _____
Techniques used to maximize response rate	• Pre-notification strategies	Which, if any, of the following pre-notification strategies were used for the face-to-face interviews that were conducted in this study? Please check all that apply. - Advance letter sent prior to initial visit - Email message sent prior to initial visit - Telephone call made prior to initial visit - Announcement in local newspaper, radio, or television - Other, specify: _______________ - None of the above	Were postal or telephone components used at any point (e.g., advance contacts)? Yes - postal →Question 41 Yes - telephone →Question 41 No →Question 42

	• Use of incentives • Specific incentive offers made to a particular group of sample members	How many different respondent incentives were initially used for the face-to-face interviews that were conducted in this study? For example, if half of the respondents were randomly assigned to receive €15 and the other half received a gift basket, this should be recorded as two incentives; or if €10 was included in the advance letter and €20 was promised upon completion of the interview, this should also be counted as two incentives. ____ different incentives used {1-10}	Were incentives offered? Yes, to respondent _____ Yes, to interviewer _____ No, neither to respondent nor to interviewer _____
	• Refusal conversion protocols • Interviewer incentives/ bonuses	Which, if any, of the following (additional) techniques were used to maximize response rate for the face-to-face interviews that were conducted in ? Please check all that apply. - Special refusal aversion or refusal conversion training sessions for interviewers - Specially designated interviewers for reluctant cases (e.g., flying experienced interviewers in to attempt difficult cases) - Persuasion letters sent to reluctant sample members - Increased or additional respondent incentives implemented after the start of data collection - Interviewer incentives/bonuses - None of the above	

Contact protocols	• Minimum number of contacts (on weekday, in the evening, on weekends) before a case is finalized • Maximum number of contacts after a cases would be finalized • Methods used to attempt to reach sample members	Was there a minimum number of attempts required before a sample element was finalized (i.e., no more attempts were made on the case) for face-to-face data collection in this study? - Yes - No	Were interviewers required to make a certain number of calls/ visits before they stopped approaching an address or household? Minimum number of calls/visits required - please write in number _____ No minimum call requirement ____

Eligibility screening	• Minimum number of attempts for screening (on weekday, in the evening, on weekends) before the case is finalized • Maximum number of attempts for screening after a cases would be finalized • Methods for refusal conversion for eligibility screening • Methods to reach sample members for the screening • Any additional techniques that were used to increase response rate for the screening to determine eligibility	What was the minimum number of attempts required before a case was finalized (i.e., no more attempts were made on the case)? If the mode of contact was not specified, please only provide the total number of attempts below. ____ minimum face-to-face attempts {ALLOW VALUE,1-40} ____ minimum telephone attempts {ALLOW VALUE,1-40} ____ total minimum attempts (face-to-face and telephone) {ALLOW VALUE,1-40}	Was substitution or replacement permitted at any stage of your selection process or during fieldwork? Yes →Question 67 No →Question 68

Use of special test or data collection besides survey questions		Besides the survey questions, did this study involve any of the following? Please check all that apply. - Tests of physical performance (e.g., walking speed, grip strength) - Tests of cognitive ability (e.g., memory tasks, word recognition) - Physical measurements (e.g., height, weight, blood pressure) - Collection of biological specimens (e.g., blood, saliva, urine) - Collection of environmental specimens (e.g., soil, dust) - Procurement of respondent permission to access and link respondent data from other sources (e.g., government records, medical records, employment records) - Other, specify: {TEXT BOX} - None of the above	
Locating sample members	• Tracking procedures - leader/coordinator of tracking - tracking manual/tracking team - training - between wave tracking efforts - steps/options used in tracking (relatives, friends, neighbors, and employers)	Were any tracking activities carried out to locate sample members in this study}? Please check all that apply. - Yes - No {SKIP TO DC223}	

Quality control	• Supervision • Back-checking		Were any interviews back-checked (e.g. supervisor checks later whether interview conducted)? Yes - please write in approximate proportion % ______ No ______

Glossary

Accuracy	The degree of closeness an estimate has to the true value.
Behavior coding	Systematic coding of the interviewer-respondent interaction in order to identify problems and sometimes to estimate the frequency of behaviors that occur during the question-answer process.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: *Do you smoke?* *Yes* ___ *No* ___
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Cluster sampling	A sampling procedure where units of the sampling frame that are similar on one or more variables (typically geographic) are organized into larger groups (i.e. clusters), and a sample of groups is selected. The selected groups contain the units to be included in the sample. The sample may include all units in the selected clusters or a sub-sample of units in each selected cluster. The ultimate purpose of this procedure is to reduce interviewer travel costs for in person studies by producing distinct groups of elements where the elements within each group area are geographically close to one another.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.

Computer assisted personal interviewing (CAPI)	A face-to-face interviewing mode in which a computer displays the questions onscreen, the interviewer reads them to the respondent, and enters the respondent's answers directly into the computer.
Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Contact attempt record	A written record of the time and outcome of each contact attempt to a sample unit.
Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Control charts	A time sequence plot with decision lines added. The decision lines are used to determine whether or not a process is in control.
Cooperation rate	The proportion of all elements interviewed of all eligible units ever contacted.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
Coverage bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value because some elements in the target population do not appear on the sampling frame.

Coversheet Electronic or printed materials associated with each element that identify information about the element, e.g., the sample address, the unique identification number associated with an element, and the interviewer to whom an element is assigned. The coversheet often also contains an introduction to the study, instructions on how to screen sample members and randomly select the respondent, and space to record the date, time, outcome, and notes for every contact attempt.

Disposition code A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).

Factual question A question that aims to collect information about things for which there is a correct answer. In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative records. Factual questions can be about a variety of things, such as figure-based facts (date, age, weight), events (pregnancy, marriage), and behaviors (smoking or media consumption).
Example: *Do you smoke?*

Fitness for intended use The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."

Focus group Small group discussions under the guidance of a moderator, often used in qualitative research that can also be used to test survey questionnaires and survey protocols.

Gross sample All eligible and ineligible elements of a sample.

Half open interval A method of updating lists of addresses by adding previously omitted units to the sample when the units are identified geographically next to a selected unit.

Hours Per Interview (HPI) A measure of study efficiency, calculated as the total number of interviewer hours spent during production (including travel, reluctance handling, listing, completing an interview, and other administrative tasks) divided by the total number of interviews.

Imputation A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.

Informant The person who supplies a list of the eligible elements within the selected unit. For example, many in-person surveys select a sample of housing units at the penultimate stage of selection. Interviewers then contact the housing unit with the aim of convincing the member of the housing unit who responded to the contact attempt to provide a list of housing unit members who are eligible for the study. The housing unit member who provides a list of all eligible housing unit members is called the informant. Informants can also be selected respondents as well, if they are eligible for the study and are chosen as the respondent during the within household stage of selection.

Interactive Voice Response (IVR) A telephone interviewing method in which respondents listen to recordings of the questions and they respond by using the keypad of the telephone or saying their answers aloud.

Interpenetrated sample assignment, interpenetration Randomized assignment of interviewers to subsamples of respondents in order to measure correlated response variance, arising from the fact that response errors of persons interviewed by the same interviewer may be correlated. Interpenetration allows researchers to disentangle the effects interviewers have on respondents from the true differences between respondents.

Interviewer falsification Intentionally departing from the designed interviewer guidelines that could result in the contamination of the data. Falsification includes: 1) Fabricating all or part of an interview—the recording of data that are not provided by a designated survey respondent, and reporting them as answers of that respondent; 2) Deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a respondent refusal as ineligible for the sample; reporting a fictitious contact attempt); 3) Deliberately miscoding the answer to a question in order to avoid follow-up questions; 4) Deliberately interviewing a nonsampled person in order to reduce effort required to complete an interview; or intentionally misrepresenting the data collection process to the survey management.

Listing	A procedure used in area probability sample designs to create a complete list of all elements or cluster of elements within a specific set of geographic boundaries.
Majority country	A country with low per capita income (the majority of countries).
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.
Minority country	A country with high per capita income (the minority of countries).
Mode	Method of data collection.
Noncontact	Sampling units that were potentially eligible but could not be reached.
Non-interview	A sample element is selected, but an interview does not take place (for example, due to noncontact, refusal, or ineligibility).
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Nonresponse bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.
Open-ended question	A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers. Example: *What is your occupation?* *Please write in the name or title of your occupation*___________

Outcome rate	A rate calculated based on the study's defined final disposition codes that reflect the outcome of specific contact attempts before the unit was finalized. Examples include response rates (the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.), cooperation rates (the proportion of all units interviewed of all eligible units ever contacted), refusal rates (the proportion of all units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units), and contact rates (the proportion of all units are reached by the survey).
Outlier	An atypical observation which does not appear to follow the distribution of the rest of a dataset.
Paradata	Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.
Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins.
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.
Probability sampling	A sampling method where each element on the sampling frame has a known, non-zero chance of selection.
Process indicator	An indicator that refers to aspects of data collection (e.g., HPI, refusal rates, etc.).

Progress indicator	An indicator that refers to aspects of reaching the goal (e.g., number of complete interviews).
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Recontact	To have someone other than the interviewer (often a supervisor) attempt to speak with the sample member after a screener or interview is conducted, in order to verify that it was completed according to the specified protocol.
Refusal rate	The proportion of all units of all potentially eligible sampling units in which a respondent sampling unit refuses to do an interview or breaks off interviews of all potentially eligible sampling units.
Reinterview	The process or action of interviewing the same respondent twice to assess reliability (simple response variance).
Reluctance aversion techniques	Techniques that can reduce reluctance to participate in potential respondents, thereby increasing the overall response rate.
Response distribution	A description of the values and frequencies associated with a particular question.
Response latency	A method of examining potential problems in responding to particular items, measured by the time between the interviewer asking a question and the response.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Sample element	A selected unit of the target population that may be eligible or ineligible.

Sample management system	A computerized and/or paper-based system used to assign and monitor sample units and record documentation for sample records (e.g., time and outcome of each contact attempt).
Sample person	A person selected from a sampling frame to participate in a particular survey.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary sampling unit (SSU)	A cluster of elements sampled at the second stage of selection.
Silent monitoring	Monitoring without the awareness of the interviewer.
Social desirability bias	A tendency for respondents to overreport desirable attributes or attitudes and underreport undesirable attributes or attitudes.
Standardized interviewing technique	An interviewing technique in which interviewers are trained to read every question exactly as worded, abstain from interpreting questions or responses, and do not offer much clarification.

Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Substitution	A technique where each nonresponding sample element from the initial sample is replaced by another element of the target population, typically not an element selected in the initial sample. Substitution increases the nonresponse rate and most likely the nonresponse bias.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Unique Identification Number	A unique number that identifies an element (e.g. serial number). That number sticks to the element through the whole survey lifecycle and is published with the public dataset. It does not contain any information about the respondents or their addresses.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Vignette	A brief story/scenario describing hypothetical situations or persons and their behaviors to which respondents are asked to react in order to allow the researcher to explore contextual influences on respondent's response formation processes.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse.

References

[1] American Association for Public Opinion Research (AAPOR). (2003). *Interviewer falsification in survey research: Current best methods for prevention, detection and repair of its effects*. Paper presented at Ann Arbor Falsification Summit on Interviewer Falsification. Retrieved March 26, 2010, from http://www.amstat.org/sections/srms/falsification.pdf

[2] Asian Barometer. (n.d.). Surveys – Survey methods. Retrieved March 26, 2010, from http://www.asianbarometer.org/newenglish/surveys/SurveyMethods.htm

[3] Axinn, W.G., Chardoul, S., Chimire, D., Gatny, H., & Barber, J. (2008). *Techniques for creating a continuous record of relationships, contraception and pregnancy: Examples from radically different social and survey research settings*. Paper presented at the international conference on survey methods in multinational, multiregional, and multicultural contexts (3MC), Berlin, Germany. Retrieved March 26, 2010, from http://www.csdiworkshop.org/pdf/3mc2008_proceedings/session_18/Axinn.pdf

[4] Axinn, W.G., Fricke, T.E., & Thornton, A. (1991). The micro-demographic community study approach: Improving survey data by integrating the ethnographic method. *Sociological Methods Research, 20*(2), 187-217.

[5] Axinn, W.G., & Pearce, L.D. (2006). *Mixed method data collection strategies*. New York, NY: Cambridge University Press.

[6] Axinn, W.G., Pearce, L.D., & Ghimire, D.J. (1999). Innovations in life history calendar applications. *Social Science Research, 28*, 243-264.

[7] Belli, R.F. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6*(4), 383-406.

[8] Berk, M. L., Mathiowetz, N. A., Ward, E. P., & White, A. A. (1987). The effect of prepaid and promised incentives: Results of a controlled experiment. *Journal of Official Statistics, 3*, 449-457.

[9] Biemer, P. P. (2004). The twelfth Morris Hansen lecture simple response variance: Then and now. *Journal of Official Statistics, 20*, 417-439.

[10] Biemer, P. P., & Forsman, G. (1992). On the quality of reinterview data with application to the current population survey. *Journal of the American Statistical Association, 87*, 915-923.

[11] Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.

[12] Broom, A. (2005). Using qualitative interviews in CAM research: A guide to study design, data collection, and data analysis. *Complementary Therapies in Medicine, 13*, 65-73.

[13] Carlson, R. O. (1958). To talk with kings. *Public Opinion Quarterly*, *22*(3), 224-229.

[14] Chikwanha, A. B. (2005). *Conducting surveys and quality control in Africa: Insights from the Afrobarometer*. Paper presented at the WAPOR/ISSC Conference, Ljubljana, Slovenia.

[15] Choldin, H. M., Kahn, A. M., & Ara, B. H. (1983). Cultural complications in fertility interviewing. In M. Bulmer & D. P. Warwick (Eds.), *Social research in developing countries* (pp. 261-274). New York, NY: John Wiley & Sons.

[16] Cialdini, R. B. (1988). *Influence: Science and practice*. Glenview, IL: Scott, Foresman.

[17] Cohen, G., & Duffy, J. C. (2002). Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics*, *18*, 13-23.

[18] Couper, M. P., Holland, L., & Groves, R. M. (1992). Developing systematic procedures for monitoring in a centralized telephone facility. *Journal of Official Statistics*, *8*, 63-76.

[19] de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, *21*, 233-255.

[20] de Leeuw, E. D., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, A. D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (pp. 41-54). New York, NY: John Wiley & Sons.

[21] de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). Mixed-mode surveys: When and why. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 299-316). New York/London: Lawrence Erlbaum Associates.

[22] European Social Survey (n.d.). *Advance letter*. Retrieved April 1, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=120&Itemid=157

[23] Forsman, G., & Schreiner, I. (1991). The design and analysis of reinterview: An overview. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 279-301). New York, NY: John Wiley &

Sons.

[24] Freedman, D.S., Thornton, A., Camburn, D., Alwin, D., & Young-Demarco, L. (1988). The life history calendar: A technique for collecting retrospective data. *Sociological Methodology, 18*, 37-68.

[25] Fricke, T.E. (2005). Taking culture seriously: Making the social survey ethnographic. In T.S. Weisner (Ed.), *Discovering successful pathways in children's development: Mixed methods in the study of childhood and family life* (pp. 185-218). Chicago, IL: University of Chicago Press.

[26] Gallup, Inc. (2007). *Gallup World Poll research design*. Retrieved May 7, 2010, from http://www.muslimwestfacts.com/mwf/File/125945/MWResearchDesign030207.pdf

[27] Gasquet, I., Falissard, B., & Ravaud, P. (2001). Impact of reminders and method of questionnaire distribution on patient response to mail-back satisfaction survey. *Journal of Clinical Epidemiology*, *54*, 1174-1180.

[28] Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646-675.

[29] Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*(4), 475-495.

[30] Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York, NY: John Wiley & Sons.

[31] Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.), (2002). *Survey nonresponse*. New York, NY: John Wiley & Sons.

[32] Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society*: *Series A (Statistics in Society), 169*, 439-457.

[33] Hargreaves, J.R., Morison, L.A., Gear, J.S.S., Makubele, M.B., Porter, J.D.H., Busza, J., Watts, C., Kim, J.C., & Pronyk, P.M. (2006). 'Hearing the voices of the poor': Assigning poverty lines on the basis of local perceptions of poverty - A quantitative analysis of qualitative data from participatory wealth rankings in rural South Africa. *World Development, 35*(2), 212-229.

[34] Harkness, J. A. (1999). In pursuit of quality: Issues for cross-national survey research. *International Journal of Social Research Methodology*, *2*, 125-140.

[35] Heath, A., Fisher, S., & Smith, S. (2005). The globalization of public

opinion research. *Annual Review of Political Science, 8,* 297-333.

[36] Howe, G. and McKay, A. (2007). Combining quantitative and qualitative methods in assessing chronic poverty: The case of Rwanda. *World Development, 35*(2), 197-211.

[37] Howell, D. (2010). *Enhancing quality and comparability in the comparative study of electoral systems.* Chapter 30 in J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B-E. Pennell, T. Smith (Eds.) *Survey methods in multicultural, multinational, and multiregional contexts.* Hoboken, NJ: John Wiley & Sons.

[38] Ikeda, K., Yamada, M., Taniguchi, N., Nishizawa, Y., & Steel, G. (2007). *Japan country report: Second wave of Asian barometer survey* (Working paper series, 32). Retrieved March 26, 2010, from http://www.asianbarometer.org/newenglish/publications/workingpapers/no.32.pdf

[39] International Telecommunication Union (ITU). (2009). *Information society statistical profiles 2009*. Retrieved from April 1, 2010, from http://www.itu.int/ITU-D/ict/index.html

[40] Jäckle, A., Roberts, C., & Lynn, P. (2006). *Telephone versus face-to-face interviewing: Mode effects on data quality and likely causes* (ISER Working Paper No. 2006-41). Colchester: University of Essex. Retrieved March 26, 2010, from http://www.iser.essex.ac.uk/publications/working-papers/iser/2006-41.pdf

[41] Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Lawrence Erlbaum.

[42] Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Heeringa, S. G., Hiripi, E., Jin, R., Pennell, B-E., Walters, E. E., Zaslavsky, A., & Zheng, H. (2004). The US National Comorbidity Survey replication (NCS-R): design and field procedures. *International Journal of Methods in Psychiatric Research, 13*, 69-92.

[43] Kessler, R. C., Ustun, T. B., & World Health Organization. (2008). *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge; New York: Geneva: Cambridge University Press; Published in collaboration with the World Health Organization.

[44] Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, *44*, 380-387.

[45] Koch, A., Blom, A. G., Stoop, I., & Kappelhof, J. (2009). Data collection quality assurance in cross-national surveys: The example of the ESS. *Methoden Daten Analysen*, 3, 219-247.

[46] Kulka, R. A., & Weeks, M. F. (1988). Towards the development of optimal calling protocols for telephone surveys: A conditional probabilities approach. *Journal of Official Statistics, 4*, 319-332.

[47] Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision* (2nd ed.). Thousand Oaks, CA: Sage Publications.

[48] Lin, I. F., & Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly*, *59*(2), 236-258.

[49] Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

[50] Massey, D.S. (1987). The ethnosurvey in theory and practice. *International Migration Review*, *21*, 1498-1522.

[51] Morse, J.M., & Niehaus, L. (2009). *Mixed method design: Principles and procedures.* Walnut Creek, CA: Left Coast Press.

[52] Mudryk, W., Burgess, M. J., & Xiao, P. (1996). Quality control of CATI operations in Statistics Canada. *Proceedings of the section on survey research methods* (pp.150-159). Alexandria, VA: American Statistical Association.

[53] Office of Management and Budget. (2006). *Standards and guidelines for statistical surveys*. Washington, DC: Office of Information and Regulatory Affairs, OMB. Retrieved January 10, 2010, from http://www.whitehouse.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf

[54] Sampson, R.J., & Laub, J.H. (1998). Integrating quantitative and qualitative data. In J.Z. Giele & G.H. Elder Jr. (Eds.), *Methods of life course research: Qualitative and quantitative approaches* (pp. 213-230). Thousand Oaks, CA: Sage Publications.

[55] Scholz, E. & Heller, M. (2009). *ISSP Study Monitoring 2007* (GESIS Technical Reports, 2009/5). Mannheim: GESIS. Retrieved March 26, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/TechnicalReport_09-5.pdf

[56] Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 163-178). New

York, NY: John Wiley & Sons.

[57] Singer, E., Hoewyk, J. V., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics, 15*, 217-230.

[58] Skjåk, K. K., & Harkness, J. A. (2003). Data collection methods. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 179-193). Hoboken, NJ: John Wiley & Sons.

[59] Stoop, I. A. L. (2005). *The hunt for the last respondent: Nonresponse in sample surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.

[60] United Nations. (2005). *Household surveys in developing and transition countries.* New York, NY: United Nations, Department of Economic and Social Affairs. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[61] van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.

[62] van den Brakel, J. A., Vis-Visschers, R., & Schmeets, J. J. G. (2006). An experiment with data collection modes and incentives in the Dutch Family and Fertility Survey for Young Moroccans and Turks. *Field Methods, 18*, 321-334.

Internet Links

[63] Afrobarometer Survey. Retrieved July 10, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[64] Asian Barometer. Retrieved July 10, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

[65] Demographic and Health Surveys. (2010). Retrieved August 17, 2010, from http://www.measuredhs.com/aboutsurveys/

[66] European Social Survey (ESS). Retrieved July 10, 2010, from http://www.europeansocialsurvey.org

[67] International Social Survey Programme. Retrieved July 10, 2010, from http://www.issp.org/

[68] Survey of Health, Ageing and Retirement in Europe (SHARE). (2010). Retrieved July 10, 2010, from http://www.share-project.org/

[69] World Values Survey. Retrieved July 10, 2010, from http://www.worldvaluessurvey.com/

Further Reading (general)

Axinn, W. G. (1989). Interviewers and data quality in a less developed setting. *Journal of Official Statistics, 5*(3), 265-280.

Axinn, W. G., & Pearce, L. D. (2006). *Mixed method data collection strategies*. New York, NY: Cambridge University Press.

Casley, D. J., & Lury, D. A. (1981). *Data collection in developing countries*. Oxford: Oxford University Press.

de Leeuw, E. D. (1992). *Data quality in mail, telephone, and face to face surveys*. Unpublished doctoral dissertation, Vrije Universiteit, Amsterdam.

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). *International handbook of survey methodology*. New York/London: Lawrence Erlbaum Associates.

Groves, R. M. (1989). *Survey errors and survey costs*. Hoboken, NJ: John Wiley & Sons.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646-675.

Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., et al. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly, 70*(5), 720-736.

Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M. Singer, E. & Tourangeau R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons

Lyberg, L. E., Biemer, P. P., Collins, M., de Leeuw, E. D., Dippo, C., Schwarz, N. et al. (Eds.). (1997). *Survey measurement and process quality*. New York, NY: John Wiley & Sons.

Saris, W. E. (1998). The effects of measurement error in cross cultural research. In J. A. Harkness (Ed.) *Cross-cultural survey equivalence* (ZUMA-Nachrichten Special series, Volume (3)) (pp. 67-87). Retrieved March 26, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf

Further Reading (integrating qualitative methods into data collection)

Barahona, C., & Levy, S. (2006). The best of both worlds: Producing national statistics using participatory methods. *World Development, 35*(2), 326-341.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York, NY: Routledge.

Denzin, N.K. (1978). *Sociological methods: A sourcebook* (2nd ed.). New York, NY: McGraw-Hill.

Farmer, T., Robinson, K., Elliott, S.J., & Eyles, J. (2006). Developing and implementing a triangulation protocol for qualitative health research. *Qualitative Health Research, 16*, 377-394.

Helitzer-Allen, D., Makhambera, M., & Wangel, A-M. (1994). Obtaining sensitive information: The need for more than focus groups. *Reproductive Health Matters, 3*, 75-82.

Hines, A.M. (1993). Linking qualitative and quantitative methods in cross-cultural survey research: Techniques from cognitive science. *American Journal of Community Psychology, 21*(6), 729-746.

Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined? *Journal of Comparative Social Work, 1*, 1-12.

Liamputtong, P. (2007). *Researching the vulnerable*. London, UK: Sage Publications.

London, A.S., Schwartz, S., & Scott, E.K. (2006). Combining quantitative and qualitative data in welfare policy evaluations in the United States. *World Development, 35*(2), 342-353.

McAdam, D. (1988). *Freedom Summer*. New York, NY: Oxford University Press.

Newman, I., & Benz, C. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.

Onwuegbuzie, A.J., & Leech, N.L. (2005). Taking the 'Q' out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms. *Quality & Quantity, 39*, 267-296.

Place, F., Adato, M., & Hebinck, P. (2006). Understanding rural poverty and investment in agriculture: An assessment of integrated quantitative and qualitative research in western Kenya. *World Development, 35(*2), 312-325.

Roberts, J., & Horney, J. (2010). The life event calendar method in criminological research. In A.R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 289-312). New York, NY: Springer.

Sieber, S.D. (1973). The integration of fieldwork and survey methods. *The American Journal of Sociology, 78*(6), 1335-1359.

Tarrow, S. (2004). Bridging the quantitative-qualitative divide. In D. Collier & H.E. Brady (Eds.), *Rethinking social inquiry: diverse tools, shared standards* (pp. 171-179). Lanham, MD: Rowman & Littlefield Publishers, Inc.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. New York, NY: Sage Publications.

Van de Vijver, F.J.R., & Chasiotis, A. (2010). Making methods meet: Mixed designs in cross-cultural research. In J.A. Harkness, M. Braun, B. Edwards, T.P. Johnson, L. Lyberg, P. Mohler, B-E. Pennell & T.W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 455-473). Hoboken, NJ: John Wiley & Sons.

XIII. Data Harmonization

Peter Granda and Emily Blasczyk

Introduction

Harmonization refers to all efforts that standardize inputs and outputs in comparative surveys.

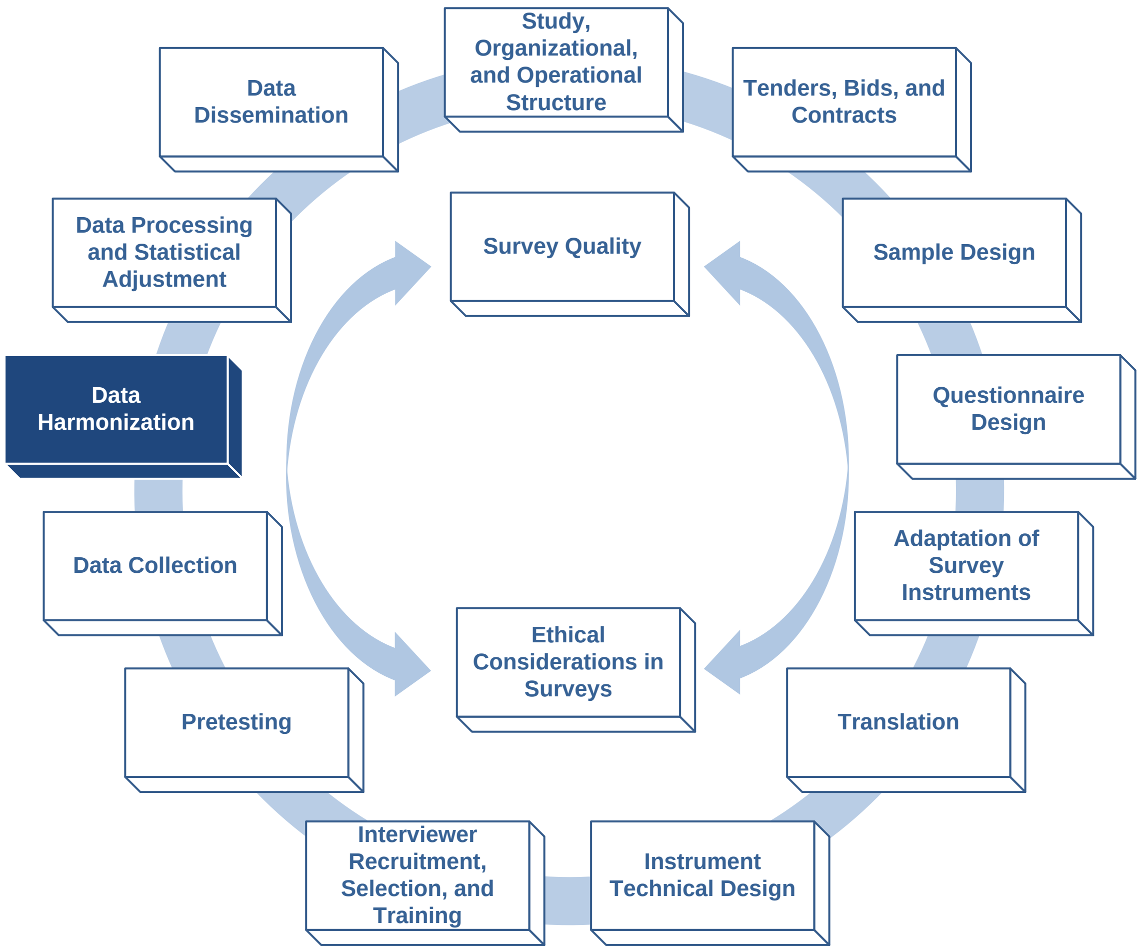
Harmonization is a generic term for procedures used predominantly in official statistics that aim at achieving, or at least improving, the comparability of different surveys and measures collected. The term is closely related to that of standardization (see Sample Design and Questionnaire Design). Harmonizing procedures may be applied in any part of the survey life cycle, such as study design, choice of indicators, question wording, translation, adaptation, questionnaire designs, sampling, field work, data coding, data editing, or documentation. The need to harmonize arises for all comparative surveys. This is particularly true if the goal is to combine the data into a single integrated dataset.

Two general approaches for harmonizing can be identified: input harmonization and output harmonization:

1. Input harmonization aims to achieve standardized measurement processes and methods in all national or regional populations. Comparability is realized through standardization of definitions, indicators, classifications, and technical requirements.
2. Output harmonization uses different national or regional measurements possibly derived from non-standardized measurement processes. These measurements are "mapped" into a unified measurement scheme. Thus, only the statistical outputs are specified, leaving it to the individual countries/regions to decide how to collect and process the data necessary to achieve the desired outputs [8] [16] [18].

Figure 1 shows data harmonization within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To ensure that survey and statistical research teams follow accepted standards when creating harmonized data and documentation files, and use a harmonization strategy that best fits their basic source materials and the objectives they wish to achieve.

1. **Decide what type of harmonization strategy to employ, taking into account that all harmonization efforts will require some combination of strategies.**

Consider "input" harmonization when the survey process is centrally coordinated.

Rationale

"Input" harmonization, usually applied in a multi-national context, seeks to impose strict standards and protocols from the beginning for the whole survey process (ex ante) by which each national survey applies the same survey procedures and a common questionnaire (see Sample Design and Translation) This strategy is meant to assure a high degree of comparability.

Procedural steps

- Provide detailed specifications, protocols, and procedures for all aspects of the survey process. The different specifications (Data Protocol, Sampling, Translation, etc.) of the European Social Survey (ESS) are a good example [4].

- Decide what to standardize.

- Create an overall monitoring team that coordinates data collection, merging national data files, and archiving.

- Publish the details of the plan and provide a schedule for the release of public-use files to the user community.

Lessons learned

- This approach involves adherence to appropriately standardized methodologies throughout the survey life cycle (see Sample Design). For example, the ESS seeks to collect data every other year, uses face-to-face interviews, aims at high-precision data, applies detailed sampling and fieldwork protocols, uses standardized translation protocols in all participating countries, aims to achieve standardized response rates, adopts consistent coding procedures, and creates and distributes well-documented datasets in a timely fashion. All of these procedures require greater organizational capabilities and resources throughout the planning and data collection stages. The results are transparent, high quality, and can produce more valuable public-use data files at the end.

- Not all comparative research will be able to follow the same procedures, so it is important to decide which methods are best, given the actual resources, survey process structure, and the intended level of precision. In addition, the creation of such common standards and their implementation at the local level requires considerable expertise. This also may not available in all multinational, multiregional, and multicultural contexts.

- It is difficult to have common standards applied once the survey is in the field. The World Mental Health Survey Initiative created such standards in the planning stage, but was unable to implement all of them, as some were not relevant in each survey country, and some countries' current survey collection practices could not support the recommended standards.

Consider "output" harmonization when the survey collection process is largely determined at the level of individual countries or cultures and there is agreement on standardization.

Rationale

This type of harmonization is implemented through two main strategies, one "ex-ante" and the other "ex-post."

Ex-ante refers to a) all measurements, such as education or employment information, which cannot be harmonized before the data collection; and b) a situation where surveys in different countries or cultures are planned to be comparable but not with the strict protocols used in input harmonization. When harmonization has already been considered during survey planning with regard to the development of common goals, measurements, and understanding of concepts, the ex-ante strategy ensures that specific targets are established for the collection of data on key variables. However, the questions used to collect these data may vary from country to country (see Questionnaire Design and Adaptation of Survey Instruments)

The second variant is an ex-post strategy, by which national statistical or survey data from archives are made comparable after the fact through a conversion procedure. Ex-post strategies can be used in situations where existing repositories will be exploited for comparative research or where intensive early planning is not possible because of financial or policy constraints.

Procedural steps

- Use an ex-ante strategy whenever possible. This enhances comparability since harmonization is addressed at the planning stage of each national data collection, as well as at the end of the process when creating harmonized data files.

- Implement an appropriate planning process.

- Use an ex-post strategy only if no consideration regarding harmonization has been given by data collectors at the start of data collection(s), but researchers later believe (e.g., because of common concepts or similar questions across surveys) that a harmonized data file can be produced through a conversion process to create comparable variables or statistics. The Integrated Fertility Survey Series is one such example [13].

- For both ex-ante and ex-post harmonization, adopt a detailed "data processing plan" that includes descriptions of how the producer(s) of the harmonized data deal with the following:
  - Differences across studies with regard to what is to be measured (e.g., definitions of population, concepts, variables).
  - Differences in how to measure (e.g., scale of measurement, wording and routing of questions, respondents).
  - Data editing.
  - Procedures used to create and define harmonized variables.
  - Construction of recoded variables (e.g., creating a common time format).
  - Sample weights and sample design variables.
  - Imputation.
  - Differences in how estimates are generated (weighting, nonresponse adjustments).

- Record all decisions about the "conversion" process systematically. One option is to use two separate databases to record all work: a production database which stores the original and harmonized materials, and a user's database which provides the analysts access to the overall process.

- Make provisions that all data conversions can be traced back to the original data.

Lessons learned

- In a working paper, Roland Gunther describes in detail the harmonization efforts surrounding the European Community Household Panel (ECHP) [9]. This survey began as a major example of input harmonization, with its design of uniform questionnaires as well as detailed definitions, rules, procedures, and models to make comparability across nations easier. After the first phase of the project, a few countries decided to cease collecting national samples for the ECHP, and instead to conduct their own national surveys, resulting in the need to do ex-post harmonization. Those doing the harmonization work learned that this kind of ex-post harmonization was resource-intensive and required staff experienced in both the original source and target formats of the ECHP framework. They also had to know in detail how their national questionnaires differed. Common problems included concepts heavily affected by national contexts, as well as differences in scales of measurement, variable coding schemes, and definitions of these concepts. Solutions to such problems were often found through ad hoc decisions about recoding, combining, or collapsing variables, and almost never through estimation techniques.

- These harmonization strategies are almost never applied exclusively on any single statistical or survey data collection. Depending on specific cultural and national characteristics, data producers should consider strategies that will enable them to collect their data in the most efficient manner. In some situations, they may want to combine strategies. For example, data producers may start with an input harmonization plan, but should be prepared to do some ex-post output harmonization to account for differences across cultures.

2. When deciding which variables to harmonize, create an initial plan and define clear objectives about what you want to achieve. The plan should include making all data conversions reversible.

Rationale

Creating a harmonization plan from the beginning of the project allows data producers to document all of their decisions at the time they are made. In case errors occur or are identified by users at a later time, all data conversions should be reversible.

Procedural steps

- Form an advisory committee of researchers knowledgeable about the subject matter at the beginning of the harmonization process, if possible, and consult with them regularly.

- Before fieldwork, consult with experts or an advisory committee on a systematic design process, and with methodology groups to investigate comparability issues.

- Show the advisory group results of the harmonization process at different points in the process to allow for possible changes in rules used to create new variables.

- Realize that not all concepts measured in the survey process are equally susceptible to harmonization efforts. For example, cross-national harmonization of the number of births and marriages is a far easier task than comparisons of divorce rates where local laws, customs, and data collection methods may differ substantially. Other concepts, such as international population migration, may not, due to a lack of precise definition and great variety in measurement criteria, lend themselves to harmonization at all, or only at the most basic level.

- Consider establishing a testing group of users knowledgeable about the subject matter but not about the harmonization process, who provide feedback on the analytic usefulness of the data before they are released publicly.

- Implement a systematic conversion creation process with appropriate quality controls.

- Identify and become familiar with software tools that facilitate a comparison of variables from different surveys, in order to determine if and how these could be harmonized. Such tools often work from a common database that stores all the information about each variable.

- Establish partnerships with producers of harmonization tools. This may be more beneficial than creating new tools, which often requires costly programming efforts.

- Where software tools are unavailable or impractical, use manual comparisons in making harmonization decisions and consult with substantive and methodological experts in doing so.

- Identify and become familiar with interactive documentation tools that allow for proper and transparent documentation.

Lessons learned

- Good decision-making about the harmonization process will benefit from the use of software tools, as well as input from a diverse group of survey researchers who can offer advice on various procedures and techniques to use when producing harmonized files. The ISSP Data Wizard [6] is used by the International Social Survey Programme (ISSP). The Data Wizard supports procedures that were previously performed manually to harmonize data at the cross-national level. The tool offers rule-based checks, automation of partial steps, and the visualization of certain conditions, to make the harmonization process more efficient, easier, and less susceptible to mistakes.

- The European Values Study (EVS) formed a number of work groups, both before and after fieldwork. The aim on the one hand was to set standards at an early stage, and on the other to consolidate and merge data which had been cleaned by participating national survey teams. This project produced an integrated source questionnaire and a set of equivalency tables to assist secondary researchers. The project web site makes all of this information easily accessible [4]. These processes and products provide critical information to secondary users of these data.

3. Focus on both the variable and survey levels in the harmonization process.

Rationale

Harmonization efforts usually concentrate on comparing and integrating information involving specific variables across data files. However, it is equally important to consider the overall characteristics of the surveys that make them good candidates for harmonization, and to report the decisions involving this process to end users.

Procedural steps

- Recognize the different aspects involved in converting source variables into target variables, such as “differences in the definitions of underlying concepts or in the definitions of the variables, deviations in the scales of measurements and so on.” The concept of citizenship, for example, presents significant challenges to researchers who want to investigate this topic [17].

- Describe similarities and differences between the source variables and the target variables, including discussion of universe statements, question wording, coding schemes, and missing data definitions.

- Consider file-level attributes when creating the harmonized data file, including how survey weights, imputation procedures, variance estimation, and key substantive and demographic concepts will change in the process.

- Pay particular attention to sampling designs and data collection methods in making assessments about the degree of comparability between different surveys.

Lessons learned

- Data producers must recognize the degrees of individual item or variable persistency when creating questionnaires and collecting data. Item persistency over time is very important in generating harmonized data files. There are considerable differences, for example, between an "absolute" persistent variable, such as "country of birth," and a less persistent variable, such as "country of citizenship." The concept might mean different things in different countries, is subject to change, and could be reported validly for multiple countries by some respondents [17].

- Quota sampling destroys comparability. Mixing surveys using quota sampling with those using probability sampling is not advisable [10] [12]. The ISSP is an example, where a comparative survey program abolished quota sampling (search for ISSP monitoring reports on the web).

- The European Social Survey (ESS) provides detailed insight into weighting issues and makes this information available. (See the ESS data site for each survey round to get the latest version).

- The Collaborative Psychiatric Epidemiology Surveys (CPES) [1] created a harmonized data file from three comparable surveys on mental health. Data producers created a pooled weight for the harmonized file, based on race/ancestry groupings and on the geographic domains of the sampling frames of each individual survey. Understanding the specific characteristics of each input file was an essential part of creating a harmonized output file [11]. All of this information was provided to users in a comprehensive explanation of the original and harmonized weights.

4. **Develop criteria for measuring the quality of the harmonization process. This includes testing it with users knowledgeable about the characteristics of the underlying surveys, the meaning of source variables, and the transformation of source variables into target variables.**

   Rationale

   Researchers may analyze harmonized files in new and unexpected ways. It is crucial to provide them sufficient information about the concepts and definitions presented, and the assumptions underlying the decisions made in their construction.

   Procedural steps

- Devise procedures to judge the quality of the harmonized outputs based on such quality criteria as consistency, completeness, and comparability.
  - Consistency can be judged by comparing the results from multiple independent efforts of harmonizing a variable; completeness is assessed based on the degree to which the original information is preserved in the harmonized data; and comparability is the degree to which the harmonized outputs can accurately report important social or economic concepts over time or between countries or cultures.
  - The Statistical Office of the European Communities (EUROSTAT) proposed the following set of quality criteria when reporting statistics which also apply to harmonization outputs (2003) [2] [15]:
    - Relevance of the statistical concepts.
    - Accuracy of the estimates.
    - Topicality and timeliness of the dissemination of results.
    - Accessibility and clarity of the information.
    - Comparability of the statistical data.
    - Coherence.
    - Completeness.
  - Strictly speaking, these traits apply to official statistical data. However, many of them would apply equally to academically produced survey data, particularly those regarding the comparability of social, economic, and demographic concepts cross-nationally or cross-culturally, and the accuracy of estimates (Total Survey Error).

- Be prepared to modify and update harmonized datasets after public release, based on comments from the research community if errors are uncovered or if certain variables need further explanation.

- Prepare presentations at social science research conferences that describe the harmonization process to potential users.

Lessons learned

- The usefulness of well-harmonized data is clearly recognized by many international organizations. For example, the United Nations Economic and Social Council indicated in a recent report that it “was working towards the harmonization of relevant environmental data-collection activities with concepts and definitions of environmental accounts. Such harmonization would result in substantial benefits in the quality of the data because it would introduce consistency checks to the environmental data and would also provide additional analytical value. The dissemination of national accounts, complemented with environment statistics information, was a very powerful analytical tool for the derivation of consistent and coherent indicators, such as resource efficiency indicators and resource use as percentage of value added. It would also allow for more in-depth analysis through scenario-modeling using input-output techniques" [19].

5. **Provide the widest range of data and documentation products about the complete process.**

Rationale

Regardless of whether researchers adopt input or output harmonization as a strategy, all aspects of the survey planning, collection, and dissemination process should be considered when producing harmonized data files or creating accompanying documentation. Users should have access not only to the harmonized end result, but also to detailed information about all steps taken by the producers, in order for them to fully understand what decisions were made during the entire process.

Procedural steps

- Define the elements of the harmonization process and start documenting it at the very beginning, in order to ensure that all decisions are captured even before a definite plan to produce a public-use data file exists.

- Document each target variable with information from all source variables, transformation algorithms, and any deviations from the intended harmonized approach, if known.

- If possible, provide users with access to the original data files used in producing the harmonized file. If direct access to original data is not permissible due to confidentiality concerns, implement procedures to assist users in proper check-backs or re-transformations. Also consider implementing some form of restricted-use data agreement to allow access under controlled conditions.

- Provide users with the code or syntax used in creating new variables for the harmonized file.

- Provide users with complete documentation, including crosswalks, which describe all the relationships between variables in individual data files with their counterparts in the harmonized file. An interactive, web-based documentation tool is often the best way to present such documentation.
  - Include original questionnaires and information about the data collection process whenever possible.

- Report on as many of the following elements of the data life cycle as apply to the particular harmonization process:
  - Project planning.
  - Sampling frame.
  - Sample size.
  - Sample design (See Instrument Technical Design, Questionnaire Design, Sample Design).
  - Extent of the field period.
  - Instrument construction and design.
  - Translation and adaptation (See Translation).
  - Mode(s) of interview.
  - Respondent follow-up if panel survey.
  - Data collection (See Data Collection, Survey Quality).
  - Editing.
  - Item nonresponse.
  - Unit nonresponse.
  - Any special treatment given to demographic and country-specific variables.
  - Sample weights.
  - Variance estimation.
  - Data production, including both long-planned and ad-hoc decisions implemented during variable conversion.
  - Documentation production.
  - Dissemination (See Data Dissemination).

  This list is based on documentation provided in the Integrated Health Interview Series (IHIS). The IHIS is an effort to provide an assortment of variables from the core household and person level files from the National Center for Health Statistics' seminal data collection effort on the health conditions for the US population from 1969 to the present. It provides extensive user notes and FAQ pages to describe how their harmonization project coped with several of these components [14].

- Consider archiving the harmonized data with a data archive to ensure continued availability of all data and documentation files and long-term preservation.

Lessons learned

- The Eurobarometer Survey Series, in operation since 1970, now includes several dozen cross-sectional surveys, all of which have been harmonized into single cross-national files before being made available to researchers. These surveys are released initially with basic information about each study and the characteristics of all variables, and are then further processed by the social science data archives, led by GESIS (German Social Sciences Infrastructure Services), to include variable frequencies, more complete documentation, and online analysis services for researchers [15]. Such partnerships between data producer and social science data archives encourage long-term preservation, enhance access, and make it possible to continually improve services to the research community.

- Some harmonization projects have gone to great lengths to describe their procedures in specific detail. For example, the Multinational Time Use Study (MTUS) has a User Guide and a comprehensive description of its coding procedures used in creating its harmonized data file [18]. Similarly, the Generations and Gender Programme (GGP) of the United Nations Economic Commission for Europe Population Activities Unit (UNECE-PAU) provides reports and guidelines about how the organization implements its harmonization decisions [7]. These projects provide transparency to both creators and users of these data and serve as an example for others to follow.

Glossary

Accuracy	The degree of closeness an estimate has to the true value.
Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population.
Anonymization	Stripping all information from a survey data file that allows to re-identify respondents (see confidentiality).
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Coding	Translating nonnumeric data into numeric fields.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Conversion process	Data processing procedures used to create harmonized variables from original input variables.

Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
Crosswalk	A description, usually presented in tabular format, of all the relationships between variables in individual data files and their counterparts in the harmonized file.
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Ex-ante	The process of creating harmonized variables at the outset of data collection, based on using the same questionnaire or agreed definitions in the harmonization process.
Ex-post	The process of creating harmonized variables from data that already exists.
Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Item nonresponse, item missing data	The absence of information on individual data items for a sample element where other data items were successfully obtained.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Mode	Method of data collection.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.

Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Probability sampling	A sampling method where each element on the sampling frame has a known, non-zero chance of selection.
Public use data file	An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quota Sampling	A non-probability sampling method that sets specific sample size quotas or target sample sizes for subclasses of the target population. The sample quotas are generally based on simple demographic characteristics, (e.g., quotas for gender, age groups and geographic region subclasses).
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Restricted-use data file	A file that includes information that can be related to specific individuals and is confidential and/or protected by law. Restricted-use data files are not required to include variables that have undergone coarsening disclosure risk edits. These files are available to researchers under controlled conditions.
Sample design	Information on the target and final sample sizes, strata definitions and the sample selection methodology.
Sample element	A selected unit of the target population that may be eligible or ineligible.

Sampling frame	A list or group of materials used identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Source variables	Original variables chosen as part of the harmonization process.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Target variables	Variables created during the harmonization process.

Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the precise estimation and reduction of the mean square error (MSE) of statistics of interest.
Transformation algorithms	Changing the values of a variable by using some mathematical operation.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Universe statement	A description of the group of respondents to which the survey item applies (e.g., "Female, ≥ 45, Now Working").
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

Sampling frame	Lists or materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. These lists or materials can include maps of areas in which the elements can be found, lists of members of a professional association and registries of addresses or persons.
Source variables	Original variables chosen as part of the harmonization process.
Strata	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Target variables	Variables created during the harmonization process.
Transformation algorithms	Changing all the values of a variable by using some mathematical operation.
Universe statement	A description of the subgroup of respondents to which the survey item applies (e.g., “Female, ≥ 45, Now Working”).
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

References

[1] Collaborative Psychiatric Epidemiology Surveys (CPES). (2010). Retrieved January 14, 2010, from http://www.icpsr.umich.edu/CPES

[2] Database of Integrated Statistical Activities (DISA). (2010). *DISA activities by statistical topic or domain: 4.1 Metadata (Eurostat)*. Retrieved April 27, 2010, from http://www1.unece.org/stat/platform/display/DISA/4.1+Metadata+(Eurostat)

[3] Eurobarometer Survey Series. (2010). Retrieved January 14, 2010, from http://www.esds.ac.uk/International/access/eurobarometer.asp

[4] European Social Survey (ESS). (2010). Retrieved May 03, 2010, from http://www.europeansocialsurvey.org

[5] European Values Study (EVS). (2010). Retrieved January 14, 2010, from http://www.europeanvalues.nl/

[6] Generations & Gender Programme (GGP). (2009). Retrieved December 17, 2009, from http://www.unece.org/pau/ggp/

[7] German Social Science Infrastructure Services (GESIS). (2010). *ISSP DataWizard*. Retrieved January 14, 2010, from http://www.gesis.org/forschung-lehre/programme-projekte/informationswissenschaften/projektuebersicht/issp-datawizard/

[8] Granda, P., Wolf, C., & Hadorn, R. (2010). Harmonizing survey sata. In J. A. Harkness, M. Braun, B. Edwards, T. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E., Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 315-332). Hoboken, NJ: John Wiley & Sons.

[9] Gunther, R. (2003). *Report on compiled information of the change from input harmonization to ex-post harmonization in national samples of the European Community Household Panel – Implications on data quality* (Working Paper #19). Retrieved January 14, 2010, from http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/Chintex/Projekt/Downloads/WorkingPaper1__092003,property=file.pdf

[10] Häder, S., & Gabler, S. (2003). Sampling and estimation. In J. A. Harkness, F. J. R. van de Vijver & P. M. Mohler (Eds.), *Cross-cultural survey methods* (pp. 117-136). Hoboken, NJ: John Wiley & Sons.

[11] Heeringa, S. G., & Berglund, P. (2007). *National Institutes of Mental Health (NIMH): Collaborative Psychiatric Epidemiology Survey*

Program (CPES) data set-- Integrated weights and sampling error codes for design-based analysis. Retrieved January 14, 2010, from http://www.icpsr.umich.edu/cocoon/cpes/using.xml?section=Weighting#I.++Introduction

[12] Heeringa, S. G., & O'Muircheartaigh, C. (2010). Sampling design for cross-cultural and cross-national studies. In J. Harkness, B. Edwards, M. Braun, T. Johnson, L. E. Lyberg, P. Mohler, B-E. Pennell, & T. Smith. (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts* (251-267). Hoboken: John Wiley & Sons.

[13] Integrated Fertility Survey Series (IFSS). (2010). Retrieved April 27, 2010, from http://www.icpsr.umich.edu/icpsrweb/IFSS/

[14] Integrated Health Interview Series (IHIS). (2010). Retrieved January 14, 2010, from http://www.ihis.us/ihis/

[15] Joint UNECE/EUROSTAT Work Session on Statistical Data Confidentiality. (2009). Retrieved March 8, 2010, from http://www.unece.org/stats/documents/2009.12.confidentiality.htm

[16] Lyberg, L. E., & Stukel, D. M. (2010). Quality assurance and quality control in cross-national comparative studies. In J. A. Harkness, M. Braun, B. Edwards, T. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E., Pennell & T. W. Smith (Eds.), Survey methods in multinational, multicultural and multiregional contexts (pp. 227-249). Hoboken, NJ: John Wiley & Sons.

[17] Minkel, H. (2004). *Report on data conversion methodology of the change from input harmonization to ex-post harmonization in national samples of the European Community Household Panel – Implications on data quality* (Working Paper #20). Retrieved January 14, 2010, from http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/Chintex/Projekt/Downloads/WorkingPaper2__012004,property=file.pdf

[18] Multinational Time Use Study (MTUS). (2009). Retrieved December 17, 2009, from http://www.timeuse.org/

[19] United Nations Economic and Social Council. (2005) *Environmental-economic accounting* (E/CN.3/2005/15). Retrieved January 14, 2010, from http://unstats.un.org/unsd/statcom/doc05/2005-15e.pdf

Further Reading

Bauer, G., Jungblut, J., Muller, W., Pollak, R., Weiss, F., & Wirth, H. (2006). *Issues in the comparative measurement of the supervisory function.* Unpublished manuscript. Retrieved May 23, 2008, from http://www.mzes.uni-mannheim.de/publications/papers/Supervisor_Function.pdf

Bilgen, I., & Scholz, E. (2007). *Cross-national harmonisation of socio-demographic variables in the International Social Survey Programme (ISSP)*. Anaheim, CA: American Association of Public Opinion Research.

Burkhauser, R. V., & Lillard, D. R. (2005). The contribution and potential of data harmonization for cross-national comparative research. *Journal of Comparative Policy Analysis, 7*(4), 313-330.

Carlson, R. O. (1958). To talk with kings. *Public Opinion Quarterly, 22*(3), 224-229.

Desrosieres, A. (2000). Measurement and its uses: Harmonization and quality in social statistics. *International Statistical Review/Revue Internationale de Statistique, 68*(2), 173-187.

Ehling, M. (2003). Harmonising data in official statistics. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (pp. 17-31). New York, NY: Kluwer Academic / Plenum Publishers.

Esteve, A., & Sobek, M. (2003). Challenges and methods of international census harmonization. *Historical Methods, 36*(2), 66-79.

Gandek, B., Alacoque, J., V, U., Andrew-Hobbs, M., & Davis, K. (2003). Translating the short-form headache impact test (HIT-6) in 27 countries: Methodological and conceptual issues. *Quality of Life Research, 12*, 975-979.

Gil Alonso, F. (2006). *Toward a European statistics system: Sources of harmonized data for population and households in Europe.* Paper presented at the International Data Session of the EAPS European Population Conference.

Hantrais, L., & Mangen, S. (Eds). (1996). *Cross-national research methods in the social sciences*. New York, NY: Pinter.

Harkness, J. A. (Ed.) (1998). *Cross-cultural survey equivalence*. Mannheim, Germany: ZUMA.

Hoffmeyer-Zlotnik, J. H. P. (2004). *Data harmonisation*. Roundtable at a conference for the Network of Economic and Social Infrastructures, Luxembourg.

Jan, W. (2009). Establishing Equivalence. In T. Landman & N. Robinson (Eds.), *The SAGE Handbook of Comparative Politics* (pp. 84-100). London: Sage

Publications.

Kallas, J., & Apostolos L. (2008). A documentation model for comparative research based on harmonization strategies. *IASSIST Quarterly*(Winter-Summer), 12-24. Retrieved May 02, 2010, from http://www.iassistdata.org/publications/iq/iq32/iqvol321_4kallas.pdf

Kennett, P. A. (2001). *Comparative social policy: Theory and research*. Buckingham: Open University Press.

Korner, T., & Meyer, I. (2005). Harmonising socio-demographic information in household surveys of official statistics: Experiences from the Federal Statistical Office Germany. In J. H.P. Hoffmeyer-Zlotnik, & J. A. Harkness (Eds.), *Methodological aspects in cross-national research.* (ZUMA-Nachrichten Spezial 11) (pp. 149-162). Mannheim, Germany: ZUMA.

Niero, M., Martin, M., Finger, T., Lucas, R., Mear, I., Wild, D., et al. (2002). A new approach to multicultural item generation in the development of two obesity-specific measures: The obesity and weight loss quality of life (OWLQOL) questionnaire and the weight related symptom measure (WRSM). *Clinical Therapeutics, 24*(4), 690-700.

Olenski, J. (2003). *SSDIS: Global standard for harmonization of Social statistics.* Unpublished manuscript. Retrieved May 23, 2008, from http://unstats.un.org/UNSD/demographic/meetings/egm/Socialstat_0503/docs/no_10.pdf

Przeworski, A. & Teune, H. (1966). Equivalence in cross-national research. *Public Opinion Quarterly, 30*(4), 551-568.

U.K. Office for National Statistics (ONS). (2008) *National statistics harmonization*. Retrieved May 23, 2008, from, http://www.ons.gov.uk/about-statistics/harmonisation/index.html

XIV. Data Processing and Statistical Adjustment

Rachel A. Orlowski, Frost Hubbard, and Emily Blasczyk

Introduction

The following guidelines detail ways in which the data collected within each country or culture must be processed (i.e., coded, captured, and edited). Although these processing steps tend to be sequential, they may also have an iterative flow. Regarding the survey lifecycle more generally, data processing does not have to wait until all the data have been collected; some of these processing steps can, and possibly should, be taken prior to or concurrent with data collection (see Instrument Technical Design and Data Collection). The flow involved in processing the survey data may also differ between paper and computer-assisted questionnaires. In computer-assisted surveys, capturing the data, performing edit checks, and building data files can, at least partially, occur automatically while the data are being collected. Some effort may be eliminated with the processing capabilities of computer-assisted interviewing. The data processing effort should be considered when determining the mode of data collection, as well as the costs associated with that decision.

After processing, the data from each country can be harmonized with those from other countries (see Data Harmonization). The calculation of outcome rates and statistical adjustments (i.e., missing value imputation, survey weight creation, and variance estimation) can be performed, as described in these guidelines. Finally, the data can be disseminated as a cross-cultural dataset (see Data Dissemination). Substantive analyses can be performed on the disseminated dataset.

Processing and adjustment activities often are not given adequate attention. This is unfortunate because costly errors can still occur after the data have been collected. Just as interviewers may introduce measurement error, data processing operators (e.g., coders, keyers) may potentially introduce processing error, sometimes systematically [4]. Often only a few errors are responsible for the majority of changes in the estimates [29]. To lessen effort, and possibly minimize error, checks can be performed throughout the field period (while respondent is still available) rather than waiting until the end of data collection. The burden of programming and checking should not be underestimated [29].

These guidelines are broken down into Data Processing Steps and Statistical Adjustment Steps. Quality and documentation guidelines are applicable to both. Please note that this chapter assumes the reader has a basic understanding of statistics and has experience in survey data management and analysis. Please refer to Further Reading or an introductory statistics textbook if a statistics refresher is needed.

Figure 1 shows data processing and statistical adjustment within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle

Guidelines

Data Processing

Goal: To code and capture data from their raw state to an edited data file that can be (1) used within the survey organization for quality assessment of the survey implementation and (2) harmonized with other countries' data files in preparation for statistical adjustment, dissemination, and eventually substantive research.

1. **Use coding to classify survey responses into categories with associated numeric values.**

Rationale

To statistically analyze survey responses, they must be transformed into numeric form; this is done by coding. Coding is both a summarization and translation process [14]. All responses to a particular survey item are summarized into a discrete number of categories. When the survey item is closed-ended (such as the response options in a "Strongly Agree—Agree—Neither Agree nor Disagree—Disagree—Strongly Disagree" scale), it is obvious how many categories are needed—five categories for a five-point scale. When the survey item is open-ended, the number of categories is not so obvious and should be determined by the analytic purpose of that survey item. Coding is also a translation process because the responses must be mapped to categories and nonnumeric category descriptions must be mapped to numeric values.

Many code structures are defined during questionnaire and instrument development, (see Instrument Technical Design); upon collecting the data, they are revisited and possibly revised. However, codes cannot be defined before data collection for some items. Respondents' responses are, instead, recorded verbatim and coded later. Data quality, in these situations, depends partly upon the interviewer recording all of the information provided by the respondent and partly upon the coder's ability to distinguish among coding categories and to assign the appropriate numeric value.

It should be noted that cross-cultural studies may have data sources other than questionnaires which require coding [14]. Such sources could include visual images/recordings, audio recordings, and samples of physical materials (e.g., dust, soil, salvia, blood, etc.).

Procedural steps

- Review survey answers for response patterns and make any necessary modifications to the precoded response options (from the instrument specifications) in order to accurately represent the range of collected data, as well as use this data review to create codes for each variable that had not been precoded. (See Instrument Technical Design for the list of instrument specifications that describe response coding.) Create code structures systematically.
  - Design the code framework with the following attributes [14]:
    - One value for each code number
    - A text label for each code number
    - A code number for each possible response category (remember to include code numbers for item-missing data—e.g., "Don't Know," "Refused," and "Not Applicable")
    - Mutually exclusive response categories for each variable

  - The appropriate number of categories to meet the analytic purpose
- With hierarchical code structures, have the first character represent the main coding category with subsequent characters representing subcategories [4]. Be especially cautious about correctly coding the first character because errors at the higher levels are usually more serious.
- Use consistent codes across survey items [14]. For example:
  - A "Strongly Agree—Agree—Neither Agree nor Disagree—Disagree—Strongly Disagree" scale would always have the values ranging from 1 = Strongly Agree to 5 = Strongly Disagree.
  - A "Yes—No" item would always have the values 1 = Yes and 5 = No (see Instrument Technical Design for an explanation of this coding convention).
  - Refused item-missing data would always have the values of 9 (or if two-digit code numbers, the values of 99—etc.).
- Keep a link from the codes to the verbatim data to facilitate quality control.
- Whenever possible and appropriate, take advantage of established coding schemes [8].

- Consider using content analysis to convert qualitative data for quantitative analysis [32].

- Determine which variables should have codes that are standardized across countries and which could have country-specific codes. This decision needs to be communicated between the coordinating center and survey organizations.

- Generate a data dictionary. There is a data dictionary entry for each survey item (see Instrument Technical Design for examples of a data dictionary entry). Each entry should contain the following information:
  - Variable ID, name, and label.
  - Data format.
  - Response options and associated code numbers.
  - Universe statements.
  - Interviewer and respondent instructions.

- Building upon the data dictionary, develop a codebook which describes how the survey responses are associated with all of the data. The codebook includes additional metadata on the survey items, such as the question text and raw frequency of responses [4].

- Decide how coding should be conducted [4].
  - For comparative surveys implement a monitoring process that guarantees comparable coding across countries.
  - Depending on resource and facility availability, as well as the data being collected, consider centralized coding (at one location, typically the survey organization) versus decentralized coding (at several locations, sometimes the coders' homes).
    - Supervisory control is easier with centralized coding. This often results in higher coder reliability.
    - Centralized coding typically involves fewer coders, with each coder having a larger workload. The larger workload can result in a higher coder design effect (see Appendix A).
    - Decentralized coding often occurs when administrative data, such as hospital records, are collected and combined into a single data source. Different hospitals and clinics may have variation in their coding procedures. It is important to consider the caliber of the various sources of data, and it should be recognized that some recoding of such data may be required [16].
  - Depending on the complexity of the questionnaire and variability among the response options, consider automated coding (where a computer program assigns code numbers) versus manual coding (where an individual assigns code numbers). Both automated and manual coding may be done on a single variable and are not necessarily mutually exclusive.
  - Consider the coding harmonization strategies needed for the data to achieve comparability across countries. For more information, see the Data Harmonization chapter.

- For automated coding, feed the responses into a computer with software that assigns appropriate code numbers based on matching the responses to a data dictionary [4].
  - Decide between using exact matching, which results in less error but also fewer assignments, or inexact matching, which has the opposite outcome.
  - Check for any responses that are left uncoded, and manually code them.
  - If using automated coding, store the code structure as a dictionary database with alternative descriptions, so a realistic response pattern can be handled.

- Properly train coders on the study's coding design, and periodically assess their abilities.

- Control manual coding by using independent verification [4].
  - Two coders code all responses separately and an additional third coder assigns code numbers where there is disagreement, as long as the third coder's assignment matches that of one of the first two coders. If all three coders disagree, a fourth coder decides the ultimate code number.
  - Independent verification is likely to result in few coding errors, especially in comparison to dependent verification (where one coder codes the data and another coder then reviews the assigned codes and changes any code number he/she deems erroneous).
    - The likelihood of two or three coders independently assigning the same erroneous code is very small.
    - Additionally, with dependent verification, obvious errors tend to be found, leaving about 50% of the errors untouched.
  - To be effective, use coders who have equivalent coding training.

- Evaluate the coding process.
  - Collect and monitor paradata on coding, such as error rates, at the variable, code number, and coder level.
  - Assess the reliability of codes.
    - A common way to calculate reliability of a code is to compute Cohen's kappa (i.e., a statistical measure that accounts for chance); kappa is most informative when there are a small number of coding categories (see Appendix B for the formula for kappa).
    - If the reliability is less than what is specified as acceptable, provide additional coder training and consider revising the coding frame.
    - Consider recoding the item if the original code is not reliable.

Lessons learned

- Although using a comprehensive data dictionary for automated coding generally results in less manual coding, expanding the dictionary does not always result in less manual coding [4]. A more descriptive data dictionary will lessen the automated coding software's ability to exactly match and assign code numbers to the responses, resulting in more manual coding. For example, the data dictionary for one of the Swedish household expenditure surveys was updated 17 times, increasing in size from 1459 to 4230 descriptions. The third update (containing 1760 descriptions) allowed 67% of the data to be automatically coded while later versions of the data dictionary could only code up to 73% of the responses—a gain of only 6% after 14 additional updates.

- Those with prior experience coding survey data may not always be the best people to code data on a particular survey. Substantive knowledge may also be necessary when selecting coders, depending on the complexity of the survey items. For example, the World Mental Health Survey employs coders who are psychologists or psychiatrists in order to diagnose verbatim responses.

- Coding errors are not trivial; they can be harmful to accurate analysis depending on type of analysis performed.

2. Capture the data into an electronic form.

Rationale

Like coding, data capture is necessary for statistical analysis. One advantage of computer-assisted questionnaires is the elimination of a separate data capture step, thus reducing the likelihood of additional processing error. When computer-assisted questionnaires are not used, keying/entry is often the first method of data capture that comes to mind. When using a paper and pencil questionnaire, it is important to capture all data provided, even when skip patterns are not followed correctly.

It is important to capture information other than the survey data, such as the information from the coversheet for each sample element, household observations, and interview details (e.g., date, time, and length of the interview); these data will aid in monitoring, evaluating, and potentially improving the data collection process.

As more advanced technology is available, there are other alternatives to data capture that essentially eliminate keying, such as optical character recognition (OCR), intelligent character recognition (ICR), mark character recognition (MCR), voice recognition entry (VRE), and touchtone data entry (TDE). ICR, commonly known as scanning, is now widely used while TDE is still not frequently used.

Procedural steps

- Determine how data capture will occur. This may vary across countries depending upon their respective amount of funding, resource availability, infrastructure constraints, and cultural feasibility. When country-specific adaptations are necessary, it is important to establish a data capture monitoring system that ensures comparability across countries.

- Use similar conventions in programming the data entry application as used when programming the survey instrument application. For example, maintain the question order and the measurement units of the survey in the data entry system (see Technical Instrument Design).

- When entering values, allow for interviewer/keyer edit checks to reduce processing error (see Guideline 3). For example, in a computer-assisted personal interview (CAPI) instrument, an age value of 233 would prompt the interviewer to confirm the value and then reenter it as perhaps 23 or 33.

- With a paper-and-pencil questionnaire, minimize the required amount of keyer judgment by having an expert, such as a supervisor, check the responses before data entry [31]. The expert should mark the questionnaire with the value to be entered when the response is not clearly indicated.
  - Review the questionnaire for [34]:
    - Illegible responses
    - Erasures
    - Markings outside the response check box
    - Crossed out (but still legible) responses
    - Added response categories (e.g., "None," "Not Applicable," "Refused," etc.)
  - If ambiguity remains, have keyers flag any concerns they have for the expert to evaluate at a later time—so as not to slow the performance of the keyers.

- Perform independent rekey verification.
  - Have two keyers work separately and then compare their work.
  - Settle discrepancies with a computer or an adjudicator [4].
  - Strive to verify 100% of the data entry [10] [14].
  - Look for the following keyer errors [34]:
    - Wrong column/field
    - Corrected/modified (misspelled) responses

- Consider automated alternatives to key entry, including [4]:
  - Optical character recognition (OCR) to read machine-generated characters.
  - Intelligent character recognition (ICR), commonly known as scanning, to interpret handwriting.
  - Mark character recognition (MCR) to detect markings (i.e., bubbles).
  - Voice recognition entry (VRE) to automatically transcribe oral responses.
  - Touchtone data entry (TDE) to interpret numbers pressed on a telephone keypad.

- When using automated systems [4]:
  - Check what was captured and manually correct any errors from misreading the raw data or omitting information (e.g., with ICR).
  - Frequently recalibrate and configure scanning equipment to minimize the frequency of with which the software misreads information (e.g., with OCR).
  - Collect paradata on the scanning operation, such as rejects and substitutes, by character and by machine.

Lessons learned

- Many established cross-cultural surveys are partly or wholly paper-and-pencil, making data entry necessary. While studies vary somewhat in the details, typically each participating country is responsible for entering and cleaning its own data, a supervisor or data manager checks questionnaires before data entry occurs, and some percentage of questionnaires is double-entered. Whatever protocol is used, it is important to fully document the data entry process.
  - Round 4 of the Afrobarometer Survey is pencil-and-paper. Each participating country is responsible for entering and cleaning its own data. Up to eight experienced data entry clerks are trained and closely supervised by a data manager who also checks completed questionnaires for mistakes before they are entered by the clerks. Data cross-checks are performed on a regular basis. Either rolling data entry or batch data entry may be employed at the discretion of the data manager. A random sample of at least 25% of all questionnaires is double-entered [35].
  - In the Asian Barometer, another pencil-and-paper survey, data cleaning involves checks for illegal and logically inconsistent values [36].
  - Round 5 of the European Social Survey (ESS) can be administered as either a pencil-and-paper or a computer assisted survey, depending upon the country's resources. National coordinators are responsible for entering and cleaning their own data and documenting their cleaning procedures before submitting the data to the ESS Archive. Files are further scrutinized for content and consistency once uploaded to the ESS Archive [38].
  - The Living Standard Measurement Study Survey (LSMS) is also pencil-and-paper and, again, each participating country is responsible for its own data editing and cleaning. Data entry operators enter the data into a specially designed program after each of the two rounds of the LSMS. Each country uses computers with specially designed software to check for accuracy, consistency, and missing data. Further data cleaning is performed by the data manager [23].

  - The World Mental Health Survey can be administered as either a pencil-and-paper or a computer assisted survey, depending upon the country's resources. Data from pencil-and-paper versions of the interview are entered manually with a data-entry program designed by the WMH Data Collection Coordination Centre. Computer assisted versions, by nature, are automated. Guidelines require all completed pencil-and-paper interviews to be edited for legibility, missing data, and reporting standards by specially trained editors. In the majority of participating countries, follow-ups are done on questionnaires with errors. Independent double entry is recommended, but keying-acceptance sampling (ranging from 10% to 20%) is accepted and used by the majority of the participating countries who evaluate keying errors. Standard coding schemes and procedures are given to all participating countries. Ten percent double coding is required. Clean datasets, checked for common errors, such as, blank or missing variables, out-of-range responses, and consistency checks, are required from all participating countries [19].

- Data entry software ranges from simple spreadsheets to sophisticated applications with built-in edit checks. If the data entry software is not universal among the participating survey organizations, then it is likely that some countries' data will be of higher quality than others'. Consider publically available software if cost is a concern. For instance, the US Census Bureau has a data entry application, Census and Survey Processing System (CSPro) [37], that is available without cost. CSPro is a software package for entering, editing, tabulating, and disseminating census or any survey data.

- Sophisticated data entry software will help the staff keying the data (for example, by accounting for skip patterns in the questionnaire). Having this level of sophistication will likely reduce entry errors but will likely cost substantially more to program and to test properly.

- Often, the same individual(s) creates many of the entry errors (often on the same variables). By limiting the number of individuals who perform data entry, it is easier to isolate potential problems and to offer appropriate follow-up training.

3. Edit the data as a final check for errors.

Rationale

Cleaning the data (e.g., correcting errors) is the primary purpose for editing, but editing can also provide information about data quality (e.g., exposing where interviewers may have difficulty performing their roles) and about improvements to future surveys (e.g., revealing where a particular design decision may be an error source) [4].

Editing can be defined as two phases: identification followed by correction, but editing can occur at various points in the survey lifecycle [4]. Incorporating editing procedures prior to and during data collection is a better allocation of resources than only after data collection. For example, in computer-assisted surveys, the application can notify the interviewers (or respondents, if self-administered) of inconsistent or implausible responses. This gives respondents a chance to review, clarify, or correct their responses. Paper surveys can include instructions telling respondents to review their responses. Prior to data capture, survey organizations can manually look for obvious errors, such as blanks. Then, during data capture, editing software can be used to check for errors at both the variable and case level.

Procedural steps

- Create editing rules that interviewers and editing staff can follow both during and after data collection [4] [14] [25] [29] [31]. Rules can include:
  - Checking for the following:
    - Wild values (e.g., out-of-range responses, unspecified response categories, etc.)
    - Imbalance values (e.g., subcategories that do not sum to the aggregate)
    - Inconsistent values (e.g., parents' ages that are not reasonably higher than their children's, males that report pregnancies, etc.)
    - Implausible outliers (e.g., extremely high or low values)
    - Multi-response variables with only one response value and single-response variables with many response values
    - Entirely blank variables
  - Comparing the current data to data from prior waves (or to data from related respondents), when applicable.
  - Verifying the correct number of digits for numeric variables.
  - Setting a minimum number of items filled to be considered a complete interview (including item-missing data on key variables).
  - Confirming the proper flow of skip patterns.
  - Flagging omitted or duplicated records.
  - Ensuring a unique identification number for every sample element, as well as a unique identification number for each interviewer.

- Establish decision rules as to whether the potential errors should be accepted as correct, changed to another value, or flagged for further investigation [4].

- Follow up on the suspicious values only if they could seriously affect the estimates, weighing the costs and logistics of recontacting the respondent [29].

- Consider using logical imputation, when appropriate.
  - Logical imputation is the process of eliminating item-missing data by reviewing data the respondent provided in prior waves or in other items within the same questionnaire and then adding the logical value.
  - For example, if a series of questions regarding the number of drinks of beer, wine, and hard alcohol consumed in the past week all have values but the final question in the series regarding the sum of drinks consumed in the past week is blank, then the total number of drinks can be logically imputed by adding the values from the individual beer, wine, and hard alcohol items.
  - This is not a statistical technique; rather values are deduced through reasoning. Be aware of the danger of creating systematic error by using such logic.

- Program computer-assisted applications to aid in the editing process during both data collection and data processing tasks.
  - Limit programming computer-assisted data capture applications to the most important edits so as not to increase the length of the survey or to disrupt the interview/entry [14].
    - Program the software to allow the interviewer/keyer to ignore the edit check (e.g., when the respondent insists on a response, when the skip pattern was not correctly followed on a paper questionnaire, etc.).
    - If the interviewer/keyer chooses to retain the original value after the edit check, program the application to allow for a comment to be written about that decision. These comments can prevent erroneous editing.
  - Editing software may not be efficient in small surveys, but it is critical in large surveys [4].

- Create a flag that indicates a change has been made to the collected data, and keep an unedited dataset in addition to the corrected dataset [25]. The latter will help decide whether the editing process adds value. If unedited data are not kept it is truly impossible to establish whether or not improvements have been made.

- Assess a random sample of each interviewer's completed questionnaires by examining the captured data. Review the use of skip patterns and the frequency of item-missing data to see if the interviewer needs additional training on navigating the instrument or probing for complete answers.

- Collect paradata on the editing process, so it can gradually improve and become less costly (see examples in Guideline 8) [4] [13].

Lessons learned

- Overediting may delay the release of the dataset, reduce its relevance to users, and be extremely expensive [4] [13]. Basically, a lot of editing is not cost-effective. Make selective editing decisions based on the importance of the sampling element or variable, the severity of the error, the costs of further investigation, and the effects of changes in the final estimates. Often, the level of detail required for any variable(s) depends strongly on the funding sources and the purpose of the estimates. These considerations should be balanced with the other needs of the study. The time and money saved by implementing selective editing can be redirected to other processing steps or other tasks of the survey lifecycle.

- Editing must be a well-organized process; if not, on-going changes to the data may actually reduce their quality [11]. Identify fields involved in the most failed edits and repair them first.

Statistical Adjustment

Goal: To improve estimates of target population parameters based on sample survey data.

4. **Use disposition codes and calculate outcome rates based on established, cited survey research standards.**

Rationale

Response rates are one indication of survey quality and can also be used to adjust survey estimates to help correct for nonresponse bias. Therefore, reporting response rates and other outcome rates based on an established survey research standard is an important part of dissemination and publication. (See Data Dissemination for more discussion on dissemination.) Additionally, outcome rates often serve as indicators of a survey organization's general performance.

Procedural steps

- Have the coordinating center provide a list of specific disposition codes and a clear description of how to code and classify all sample elements during the field period (using temporary disposition codes) and at the end of the field period (using final disposition codes). These disposition codes will allow the standardization of outcome rate calculations across countries.

- Generally, disposition codes identify elements as a completed interview or non-interview. Non-interviews are further subdivided depending upon whether the sample element is eligible or ineligible to participate in the study. For surveys where sample elements are people, ineligible non-interviews might include the respondent being deceased, the housing unit being unoccupied, or the respondent having emigrated outside of the boundaries of the study area. Eligible non-interviews include refusal to participate, noncontacts, and other (defined by study).
- Disposition codes are mutually exclusive, and, while each sample element may be assigned different temporary disposition codes across the field period, ultimately it will be assigned *only one* final disposition code.

- Based on an established survey research standard, assign all sample elements into mutually exclusive and exhaustive categories and calculate response rates.
  - Assigning elements into predetermined final categories, those necessary in calculating a response rate, makes it possible to recalculate each country's response rate in a standard way for comparison across countries, as appropriate.
  - The World Association for Public Opinion Research/American Association for Public Opinion Research (WAPOR/AAPOR) provides one example of an established survey research standard [1].
    - According to WAPOR/AAPOR's "Standard Definitions of Final Dispositions of Case Codes and Outcome Rates for Surveys," there are four main response rate components. These are Interviews—and three categories of Non-interviews: Non-interviews-Eligible, Non-interviews-Unknown Eligibility, and Non-interviews-Ineligible.
    - WAPOR/AAPOR defines six separate response rates (RR1-RR6) [1].
      - Response rates ending in odd numbers (i.e., RR1, RR3, and RR5) do not consider partially-completed interviews to be interviews. Response rates ending in even numbers (i.e., RR2, RR4, and RR6) consider partially-completed interviews to be interviews.
      - RR1 and RR2 assume that all sample elements of unknown eligibility are eligible.
      - RR3 and RR4 estimate the percentage of elements of unknown eligibility that are actually eligible.
      - RR5 and RR6 assume that all elements of unknown eligibility are ineligible.

- The Tenders, Bids, and Contracts appendices F, G, and H contain templates that could be used to define the different response rate components and record counts of the different components.

- Based on an established survey research standard, calculate other important outcome rates such as contact, cooperation, or refusal rates.
  - There are many different industry standards available. WAPOR/AAPOR's outcome rate calculations are an example of one such standard [1]. Another has been developed by Statistics Canada [30].

Lessons learned

- Ensure that each disposition code is clearly described and reviewed during each participating country's study training. Countries may not be familiar with the specified disposition codes or the response rate terminologies. As another check, consider obtaining contact attempt records from each country early in the data collection period in order to ensure that all countries are correctly identifying different outcomes and understand the difference between temporary and final disposition codes. Implement all disposition codes according to the study requirements.

- Standardize the specific disposition codes as much as possible across all participating countries. However, recognize that some special, country-specific disposition codes may need to be created to adequately describe the situation. For example, since best practice suggests allowing the sample design to differ across countries, different disposition codes regarding ineligible elements may need to be created for certain countries.

5. Develop survey weights for each interviewed element on the sampling frame.

Rationale

Depending upon the quality of the sampling frame, the sample design, and patterns of unit nonresponse, the distribution among groups of observations in a survey dataset may be much different from the distribution in the survey population. To help correct for these differences, sampling statisticians create weights to reduce the sampling bias of the estimates and to compensate for noncoverage and unit nonresponse. An overall survey weight for each interviewed element typically contains three adjustments: 1) a base weight to adjust for unequal probabilities of selection (w_{base}); 2) an adjustment for sample nonresponse (adj_{nr}); and 3) a poststratification adjustment (adj_{ps}) for the difference between the weighted sample distribution and population distribution on variables that are considered to be related to key outcomes. If all three adjustments are needed, the overall weight is the product of these three adjustments:

$$w = w_{base} * adj_{nr} * adj_{ps}$$

However, it is not always necessary to create all three weight adjustments when creating an overall survey weight. Create the adjustments only as needed. For example, if all elements had equal probabilities of selection, a base weight would not be necessary. The overall survey weight would then be the product of any nonresponse adjustment and any poststratification adjustment [14].

Presently, the field of survey research lacks a methodology that can help develop weights for other major survey errors, such as processing and measurement error. At this time, evaluation methods are used instead of developing weights.

Procedural steps

- If necessary, calculate the base weight for each element.
  - Each element's base weight is the inverse of the probability of the selection of the specified element across all stages of selection.

- If necessary, calculate the nonresponse adjustment for each element.
  - There are many ways to calculate nonresponse adjustments. This guideline will only explain one method that uses observed response rates within selected subgroups. This method is easier to calculate than others but assumes that all members within a specific subgroup have the same propensity of responding. For information on other nonresponse adjustment methods, see [2] [28]
  - Compute response rates for mutually exclusive and exhaustive subgroups in the sample that are related to the statistic of interest.
  - The inverse of a subgroup's response rate is the nonresponse weight for each eligible, sampled element in the subgroup.

- If necessary, calculate the poststratification adjustment.
  - Multiply $w_{base} * adj_{nr}$ to obtain a weight that adjusts for both unequal selection probabilities and sample nonresponse for each eligible element.
  - Using this weight, calculate a weighted sample distribution for certain variables related to the statistics of interest where the population distribution is known (e.g., race and sex). See [18] for a method of computing poststratification weights when the population distribution is unknown for certain subgroups (e.g., using raking or iterative proportional fitting).

- In comparative surveys, make sure that the official statistics used by each participating country to estimate the population distribution have the same level of accuracy. If that is not the case, seek corrections or alternatives.
- Divide the known population count or proportion in each poststratum by the weighted sample count or proportion to compute adj_{ps}.
    - For example: According to 2007 estimates from Statistics South Africa, women comprised 52.2% of the total population residing in the Eastern Cape Province. Imagine the weighted estimate of the proportion of women in the Eastern Cape from a small local survey after nonresponse adjustments was 54.8%. The poststratification adjustment, adj_{ps}, for female respondents in the Eastern Cape would be .522/.548 = .953.
- Note that missing values for any variable needed for poststratification adjustments should be imputed (see Guideline 6 for information on imputation).

- Multiply the needed weight adjustments together to determine an overall weight for each element on the data file.

- Trim the weights to reduce sampling variance.
    - Survey statisticians trim weights by limiting the range of the weights to specified upper and lower bounds (e.g., using no less than the 10^{th} percentile and no more than the 90^{th} percentile of the original weight distribution).
    - Trimming of weights produces a reduction in sampling variance but might increase the mean square error [3].

- If necessary, consider other weight components (besides the base weight, nonresponse adjustment, and poststratification adjustment).
    - There may be weight components other than the three described in this guideline. Other possible weight components are country-specific adjustments and weights that account for differential probability of selection for certain questionnaire sections.

- Apply the final weight to each record when calculating the statistic of interest.

- Understand the advantages and disadvantages of weighting.
    - Advantages:
        - Weighting can reduce coverage bias, nonresponse bias, and sampling bias at the country or study level, depending on whether the weights were designed to reflect the population of a specific country or the entire study.
    - Disadvantages:

- Weighting can increase sampling variance. See Appendix C for a rudimentary measure of the increase in sampling variance due to weighting.
- When forming nonresponse adjustment classes, it is assumed that respondents and nonrespondents in the same adjustment class are similar. This is a relatively strong assumption.
- If the accuracy of the official statistics used to create poststratification adjustments differs by country, comparability across countries can be hampered [12]. In addition, if the poststratification adjustments do not dramatically impact the survey estimates, consider not using the adjustment.

Lessons learned

- Ensure that all participating countries thoroughly document their sampling procedures and selection probabilities at every stage of selection. Countries that do not routinely employ survey weights or use complex survey designs may not be accustomed to recording and maintaining this information. Without this information, it can be very difficult to recreate base weights once data collection is complete.

6. Consider using single or multiple imputation to compensate for item-missing data.

Rationale

Item-missing data are common in social science research data. Imputation is often used to address this problem. The aim of imputation is to reduce the bias in the estimate of the statistic of interest caused by item-missing data and to produce a rectangular dataset without holes from the missing data that can be analyzed by standard software.

The two main methods of imputation—single and multiple imputation—are described in this guideline [17] [24].

Single Imputation Methods

Rationale

Single imputation involves replacing each missing item with a single value based on the distribution of the non-missing data or using auxiliary data. [33]. It is the easier of the two imputation methods, and those less experienced and knowledgeable about imputation will likely be able to correctly implement single imputation. There are several common methods, which are discussed below.

Procedural steps

- Select a single imputation method. Consider the following:
  - Overall mean value hot-deck imputation.
    - Replace the missing values for a variable with the mean value for that variable across the entire dataset.
    - While this is a very simple method to use, it can distort the distribution of the variable with imputed values by creating a spike in the distribution at the mean value, potentially biasing the results.
  - Overall mean value cold-deck imputation.
    - Replace the missing values for a variable with the mean value for that variable from an external source or dataset.
  - Sequential hot-deck imputation.
    - Sort the dataset by specific, observed variables related to the statistic of interest. For example, imagine the statistic of interest is the average, yearly personal income in Spain. Assume that it is known from previous studies that the yearly personal income in Spain is related to years of education and age. The dataset would first be sorted by years of formal education and then respondent age.
    - See if the first element on the sorted dataset has a value for the variable that is to be imputed; in the above example it would be reported yearly personal income.
    - If the first element does not have a value, impute the mean value of the variable based on the sample elements with data on the statistic of interest.
    - If the first element does have a value, keep this reported value and move to the second element. The last reported value is now the "hot-deck" value.
    - If the second element is missing a value for the specified variable, impute the "hot-deck" value. The value for the second element then becomes the "hot-deck" value for the third element, etc.
    - Sequential hot-deck imputation is less costly than regression imputation methods because no model fitting is necessary, and it has fewer complexities than regression imputation methods. Thus, sequential hot-deck imputation is more easily understood by analysts and can reduce variance and nonresponse bias. One disadvantage is that one record may be a donor multiple times in a way that is difficult to control.
  - Regression imputation.
    - Carefully create a regression model for a specific variable that predicts the value of the variable based on other observed variables in the dataset. For example, one could create a

regression model that predicts the number of doctor visits in the past year based on demographics, such as age, sex, race, education, and occupation.
- Check that the predictor variables do not have many missing values.
- Regression imputation can produce better imputations of missing values than hot-deck methods for variables with complex missing data patterns and for small samples.

- For all variables for which at least one value was imputed, create imputation flag fields that indicate which items for each record on the data file were imputed.

Lessons learned

- The Standard Cross-Cultural Sample (SCCS) researchers who employ case deletion are frequently forced to collapse regions together in order to have enough cases to analyze. By imputing data, regional distinctions can be maintained [6].

- Sampling statisticians advise users to avoid imputing attitudinal variables since attitudes can easily change over time and missing data patterns can be difficult, if not impossible, to predict. Imputation models for factual variables are generally easier to specify because they are more static and outside validation can be provided.

- If item nonresponse is missing at random (MAR) given the covariates used in the imputation process, imputation reduces bias, sometimes a lot. In MAR, the process causing missing values can be explained either by the variables in the model or by variables from auxiliary data. (See Appendix D for more information about assumptions for missing data).

- Imputed data are fabricated data. Variances using single-imputed data methods are smaller than the true variances.

- Data analysts must be able to identify real values and imputed values. Therefore the imputation must be thoroughly documented.

Multiple Imputation Methods [22] [27]

Rationale

The goal of multiple imputation is to account for the decreased uncertainty imputed values have compared to observed values. Multiply imputed values and multiple datasets are created for each missing value. Variation in the estimates across the trial runs allows for the estimation of both sampling and imputation variance. Therefore, multiple imputation creates a distribution of imputed values that have their own standard errors and confidence intervals [33]. As previously noted, an added level of expertise is needed to perform multiple imputation, which may result in a more expensive procedure than using single imputation.

Due to the statistical complexity of multiple imputation methods, only the most commonly used method is briefly described below. Please refer to [21] for information on other methods.

Procedural steps

- Select a multiple imputation method; consider sequential regression imputation.
  - Create multiple datasets where each missing element is based on a different trial run of a regression model for each imputed item.
    - This is an iterative process where one item is imputed using an imputation model and then the next item is imputed with a regression model that uses the imputed values of the first item.
    - Consider using the same set of variables for all imputations to reduce the risk of over-fitting the model.
  - Several statistical software packages are capable of multiple imputation. IVEWare, a package developed at the University of Michigan and available to users without cost, is an example of one such package [39]. R programs that perform multiple imputation are also available [7].
  - Use sequential regression imputation when records contain different numbers of missing items.
  - Although sequential regression imputation accounts for the increased uncertainty of imputed values, it can be time-consuming for large surveys.

Lessons learned

- Even with the continual improvements in statistical software, multiple imputation methods may be hard to do for all cross-cultural surveys because it takes a greater skill level and often more time and money than single imputation. In addition, each variable requires specific treatments and evaluation on how to impute the missing values.

- Consider checking that the imputation model fits the data correctly and is well specified. A poor imputation model can actually increase the bias of the estimate, making it worse than not using imputation.

7. When calculating the sampling variance of a complex survey design, use a statistical software package with the appropriate procedures and commands to account for the complex features of the sample design.

Rationale

The survey sample design determines the level of precision. Unfortunately, many statistical texts only discuss the sampling variance formula for simple random sampling without replacement (a sampling method that is almost never used in practice). Similarly, statistical software packages (e.g., STATA, SAS, and SPSS) assume simple random sampling without replacement unless otherwise specified by the user. However, compared to a simple random sample design, stratification generally decreases sampling variance while clustering increases it (see Sample Design for in-depth explanations of simple random samples, stratification, and clustering). If the correct formulas or appropriate statistical software procedures and commands are not applied, the calculation of the precision (i.e., sampling variance) of the statistic(s) of interest can be underestimated or overestimated. Therefore, analysts are cautioned to ensure they are applying the correct methods to calculate sampling variance, based on the sampling design. Always compare results with the default simple random sample selection assumptions to check for inconsistencies that might occur due to defective estimators.

Procedural steps

- In order to use Taylor series variance estimation, which many statistical software packages use as a default, the survey data file must include, at a minimum, a final survey weight, a stratum identifier, and a sampling unit identifier for each responding sample element [14]. The chosen statistical software package must have the capacity to account for survey weights, stratification, and sampling units in the estimation process [21] [5].
  - If the complex survey design used clustering, the survey data should also include cluster identifiers for each responding sample element.
  - In order to estimate the sampling variance within a stratum, at least two selections must be made within the stratum. For a sampling design that selects only one primary sampling unit (PSU) per stratum, the sampling variance cannot be estimated without bias. In "one PSU per stratum" designs, the PSUs are arranged after data collection into a set of sampling error computational units (SECUs) that can be grouped into pairs for purposes of estimating approximate variances. If a participating country uses a sample design that selects only one PSU per stratum, the survey data must

include the SECU of each element to make variance estimation possible.

- When a survey data file is supplied with a series of replicate weights plus the final survey weight, balanced repeated replication or jackknife repeated replication could be used to estimate variances (see Appendix E).

- When estimating means and variances with statistical software packages, use the appropriate procedures and commands to account for the complex survey data. For example, SAS version 9.1.3 features the SURVEYFREQ and SURVEYMEANS procedures with strata and cluster commands to account for complex survey designs.

Lessons learned

- Not all countries may have access to statistical software packages. Therefore, it may be necessary to arrange for reduced fees or for centralized analysis. Alternatively, consider using free, open source software, such as R.

Data Processing and Statistical Adjustment

8. **Implement quality checks at each stage of the data processing and statistical adjustment processes.**

Rationale

Ensuring quality is vital throughout the survey lifecycle. Even after data collection is complete, the survey organization must continue to implement quality measures to help reduce or eliminate any errors that could arise during the processing and adjustment procedures discussed above. If the emphasis on quality is relaxed during these latter activities, all of the time and money spent on maintaining quality during the previous tasks of the survey lifecycle will be compromised.

Procedural steps

- Continually monitor coding activities, such as the number of responses that were coded automatically; were changed after data dictionary updates; and were coded in error due to coding mode, category, or data dictionary updates [4].

- Use data entry tools to perform keying quality checks. Have human analysts check for representativeness and outliers [31].

- Monitor editing using some key process statistics [4] [13]. Examples are as follows (where objects can refer to fields, characters, or records):
  - *Edit failure rate = # of objects with edit failures / # of objects edited* (estimate of amount of verification).
  - *Recontact rate = # of recontacts / # of objects edited* (estimate of number of recontacts).
  - *Correction rate = # of objects corrected / # of objects edited* (estimate of the effect of the editing process).

- Remove any identifying information from the production data. For example, remove any names and addresses attached to each responding element or unit. (For more information, see Ethical Considerations in Surveys.)

- When possible, use auxiliary data (e.g., census or population files) and paradata for post-survey adjustments and to enhance the precision of the survey estimates. For example, population files could be used to create nonresponse weighting adjustment categories. However, in cross-cultural surveys be aware of very different levels of accuracy across countries for such information.

- Compare the sum of the base weights of the initially sampled elements to the count *N* of units on the sampling frame. If the sample was selected with probabilities proportional to size, then the sum of base weights is an estimate of *N.* If an equal probability sample was selected within strata or overall, then the sum of base weights should be exactly equal to *N*.

- Assign a second sampling statistician to check the post-survey adjustment methodology and the statistical software syntax of the survey's primary sampling statistician. This should be done whether the statistical adjustments are done individually by each participating country or done for all countries by a statistical team selected by the coordinating center.

Lessons learned

- Make certain that all identifying information is removed from the dataset before making it publicly available. In some surveys, this may require detailed geographic identifiers be removed. One survey publicly released a dataset that included variables which made it easy to personally identify each respondent. The principles of the Helsinki Declaration should be upheld (see Ethical Considerations in Surveys) [15].

- When using official statistics for poststratification adjustments, consider the reputation of the agency. It has been suggested that some countries have manipulated official statistics. Examples of potential manipulations include the adjustment of agricultural outputs or redefining terms such as unemployment [9] [26].

9. Document the steps taken in data processing and statistical adjustment.

Rationale

Over the course of many years, various researchers may wish to analyze the same survey dataset. In order to provide these different users with a clear sense of how and why the data were collected, it is critical that all properties of the dataset be documented.

Documentation will help secondary data users better understand post-survey statistical adjustments that can become quite intricate, such as the imputation procedures and the creation of survey weights for complex survey designs. A better understanding of these adjustments will help ensure that secondary data users correctly interpret the data. In addition, post-survey documentation will indicate whether the survey organization that conducted the survey met benchmarks agreed to in the contract by the coordinating center and the survey organization.

Procedural steps

- Document the procedures and quality indicators of the data processing. Examples include:
  - Data capture process.
  - Versions of the data dictionary and codebook.
  - Training protocol and manuals for data coding, entry, and editing.
  - Who coded, entered, and edited the data.
  - Evaluation protocol for data coding, entry, and editing.
  - What items were coded or recoded.
  - Measure of coding reliability (e.g., Cohen's kappa).
  - Verification protocol for coding and entry.
  - Data entry accuracy rate.
  - Protocol for editing open-ended responses (e.g., remove identifying information, correct typographical errors, standardize language).
  - How the raw data were corrected during the editing process.

- Describe how the sample identification numbers/codes were assigned to each sampling unit.

        - For documentation and dissemination create, a unique identification number. This number contains no information about responding units; it is simply a unique identifier.
        - Sampling frame information should be added for internal use **only** as separate variables systematically (e.g., country two digits (06), area segment three digits (005), sample replicate three digits (002), household three digits (001), respondent selected two digits (01), etc.
        - For internal documentation, describe how the unique sample identification number/code was assigned for internal use datasets (e.g., 0600500200101: first two digits identify the country, the next three digits identify the area segment, the next three digits identify the sample replicate, the next three digits identify the household, the final two digits indicate the order of selection of the respondents within the unit where 01=main respondent selected and 02=second respondent selected).
        - For public use datasets, use a different sample identification number to prevent disclosing a respondent's identity (see Ethical Considerations in Surveys).

    - If survey weights were generated for the study, clearly explain how each individual weight adjustment was developed and how the final adjustment weight was calculated.
        - Each explanation should include both a written description and the formula used to calculate the weighting adjustment. Below are examples of the first sentence of an explanation for different weight adjustments. These are not meant to be exhaustive explanations, and the documentation of each adjustment should include further written descriptions and formulas.
            - The base weight accounted for oversampling in the Wallonia region (Belgium) strata.
            - The nonresponse adjustment was the inverse of response rate in each of three regions – Vlanders, Wallonia, and Brussels.
            - The poststratification adjustment factor adjusted weighted survey counts to totals from Denmark's 2003 population register by sex, education, and age.
            - As of March 1, 2004, a random half of the outstanding elements in the field were retained for additional follow-up efforts, and this subsample of elements was given an extra weight adjustment factor of $W = 1/.5 = 2.0$.
        - If additional adjustments were used to calculate a final weight, provide a clear description of how these components were created. Examples of additional weight components are country-specific adjustments and adjustments that account for differential probability of selection for certain questionnaire sections.

- Address whether there was any trimming of the weights and, if so, the process used to trim the weights.
- Address whether a procedure was used for scaling of the weights (e.g., population (N), population (N in 1000s), sample size (centered)).
- If a replicated weighting method was used (i.e., Jackknife Repeated Replication or Balanced Repeated Replication – see Appendix E), provide the replicate weights for variance estimation.
- Clearly describe when each of the survey weights and adjustments that were developed for the study should be used in data analysis.

- If values were imputed for specific variables in the study, clearly describe the imputation method that was used in the post processing documentation. In addition, for each variable where at least one value was imputed, create an imputation indicator variable that identifies whether a value was imputed for the specific variable or record in the dataset.

- For complex survey data, identify the cluster and stratum assignment variables made available for sampling error calculations. For instance:
  - The variable that identifies the stratum to which each sample element and sample unit belongs.
  - The variable that identifies the sampling cluster to which each sample element and sample unit belongs.
    - If the sample design has multiple stages of selection, document the variables that identify each unique sample element's primary sampling unit (PSU), secondary sampling unit (SSU), etc.
    - If Balanced Repeated Replication variance estimation was used, identify the stratum-specific half sample variable, i.e., a field that identifies whether a unit is in the sampling error computation unit (SECU) 1 or 2.

- If the risk of disclosing respondent identities is low, consider providing the different weight components on public use datasets. However, preventing disclosure of respondent identity takes priority over providing weight components.

- Discuss whether the survey met the requirements (e.g., response rates, number of interviews) outlined in the contract.
  - If the requirements were not met, provide possible reasons why the survey failed to meet these requirements.

Lessons learned

- The application of a unique identification code is often underestimated by survey agencies using their internal reference systems. For instance, a European survey implemented a two-year special panel survey where the agency conducting the study did not understand the need to link the two panel waves via one variable. Hence, the agency provided a set of hard to interpret 'synthetic' codes that made it difficult for users to know if they were correctly analyzing the data. Much time and money were spent disentangling these codes and clarifying dubious cases.

- Secondary users of survey data often have a hard time understanding when and if they should use weights in their analyses. This issue is exacerbated in many cross-cultural surveys, where participating countries may apply different nonresponse and postratification adjustment strategies. Without a clear documentation of how each country created their survey weights and when to use each of the weights in data analysis, the chance of secondary users either not applying or incorrectly applying weights and producing estimates that do not accurately reflect the respective target population greatly increases. Therefore, clear and understandable documentation of the statistical adjustment processes is extremely important.

Appendix A

Coder design effect [14]

The coder design effect ($deff_c$) applies much of the same logic as interviewer design effect (see Interviewer Recruitment, Selection, and Training chapter).

- In the formula for coder design effect below, ρ_c is the intraclass coefficient for coders, m is the average number of cases coded per coder, and r is the reliability of a particular (item) code.

$$deff_c = 1 + \rho_c (m\text{-}1)(1\text{-}r)$$

- The intraclass coefficient for coders is a measure of the ratio of coder variance to the total variance and is defined as:

$$\rho_c = \frac{(between\text{-}coder\ variance)}{(between\text{-}coder\ variance) + (within\text{-}coder\ variance)}$$

Appendix B

Cohen's kappa

Cohen's kappa can be used to assess the reliability of an individual (item) code.

- In the formula for kappa below, Pr(a) is the relative observed agreement among coders for a given item, and Pr(e) is the hypothetical probability of chance agreement in the observed data calculated from the probabilities of each coder randomly reporting each possible code category for that item.

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

- If the coders are in complete agreement then kappa equals 1. If there is no agreement among the coders (other than what would be expected by chance) then kappa is less than or equal to 0.
- Kappa values between 0.7 – 0.8 are considered reliable.

Appendix C

Loss in precision of estimate due to weighting in household surveys

While overall survey weights help decrease three different sources of bias (coverage, nonresponse, and sampling), the variability of the weights also can increase the sampling variance in household surveys. The formula below is a simple model to measure the loss in precision (L_w) due to weighting. It assumes that the weights and the variable of interest are not related.

$$L_w = \left[\frac{\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i \right)^2} \right] (n) - 1$$

- For example, if $L_w = .156$, then the sampling variance of the estimate increased by 15.6% due to differential weighting.
- L_w can also be calculated for subgroups.
- N.B.: This formula does not apply to surveys of institutions or business establishments where differential weighting can be efficient.
- This is only one method for measuring the variability of the weights.

Appendix D

Assumptions of missing data [4]

The differences between the three missing data mechanisms depend on the relationship of the variable of interest with missing observations and the variables available to explain the missingness.

- Missing Completely at Random (MCAR)
  - This missing data mechanism assumes the underlying process causing missing data is uncorrelated with any of the variables in the dataset. In other words, the probability of an observation for variable y being missing does not depend on measurements (x or y in the diagram below) in the dataset itself. An example of MCAR data is missing data due to an instrument malfunction. If MCAR holds, listwise deletion (i.e., an entire record is excluded from analysis if any one value is missing) can be employed because the available cases constitute a random subsample. Therefore, under MCAR, valid inferences to the target population can be made when analyzing only those units with complete data. If there are variables in the dataset (x, y) that help predict the missing values, the assumption does not hold. MCAR rarely holds, and, thus, listwise deletion will seldom be appropriate.
  - The concept of MCAR is illustrated below where y is the variable of interest with missing values, x is a predictor of y, m is the process causing missingness, and q is a variable not in the dataset.

$$x \longrightarrow y$$

$$q \longrightarrow m$$

- Missing at Random (MAR)
  - MAR is a weaker assumption about missingness than MCAR. In MAR, the process causing missing values can be explained by observed, non-missing data (x in the diagram below) other than the variable of interest y. Said another way, the probability of data missing on variable y is not related to the value of y, controlling for other variables. For data that are MCAR or MAR, the missing data mechanism is deemed ignorable. Note that the missing data *mechanism* is what is ignorable, not the missing data themselves. For data that are MAR, imputation will reduce bias.
  - The concept of MAR is illustrated below where y is the variable of interest with missing values, x is a predictor of y and also can

predict the mechanism for missing values, m. q is auxiliary to the dataset and also predicts m.

$$\begin{tikzcd} x \arrow[r] \arrow[dr] & y \\ q \arrow[r] & m \end{tikzcd}$$

- Missing Not at Random (MNAR)
  - For data that are MNAR, even after controlling for other observed variables in the dataset (x in the diagram below), the reason for a variable y having missing observations still depends on the unseen observations of y itself. One example of data that could be MNAR is reported income. Individuals with either high or low incomes can be reluctant to report how much they earn. If this is true, the probability of obtaining a measure of a person's income will depend upon the amount the person earns. Nonignorable nonresponse creates data that are MNAR, and, hence, a method of imputation that accounts for this is necessary.
  - In the diagram below, y is the variable of interest with missing values, x is a predictor of y in the dataset, and q is unobserved auxiliary data. The three variables y, x, and q all predict m, the mechanism of missing values.

$$\begin{tikzcd} x \arrow[r] \arrow[dr] & y \arrow[d] \\ q \arrow[r] & m \end{tikzcd}$$

Appendix E

Estimating complex statistics when sample size is not fixed

Whenever the sample size is not fixed, use the Taylor Series estimation or one of the replicated methods, such as Balanced Repeated Replication (BRR) or Jackknife Repeated Replication (JRR), to estimate ratio means or other complex statistics.

- Taylor Series estimation.
  - This method computes the sampling variance of an approximation to a complex function like a ratio or regression coefficient. (See [20] for the exact formulas.)
  - Advantages:
    - Used by most statistical software packages.
  - Disadvantages:
    - Requires analytic manipulations and computation of derivatives (but these have been done by developers of the software packages for common type of estimates).
    - Not useful if estimate cannot be expressed as a function of sample totals.
    - Taylor Series estimates in most software packages do not account for the variability of nonresponse adjustments.

- Balanced Repeated Replication (or Half-Sample Replication).
  - This method assumes a paired selection design (i.e., 2 PSUs per stratum) and selects *H** half sample replicates (*H** is the smallest multiple of 4 greater than or equal to the number of strata) by deleting one primary sampling unit (PSU) from each stratum according to the pattern in a Hadamard matrix. Each remaining element in the half sample receives a replicate weight of two. Fay's method of BRR is an alternative that retains both PSUs in a pair but modifies their survey weights [17].
  - Advantages:
    - More useful for complex estimates, such as medians, than Taylor Series.
    - Easily applied to user-specified statistics like differences or ratios of domain means.
    - Accounts for variability due to multiple steps in adjustment more easily than does Taylor Series.
  - Disadvantages:
    - Best used only with a paired selection stratification design.
    - Appending replicate weights to each record increases file size.

- Combining of strata and PSUs is sometimes done to reduce number of replicates. This must be done carefully to avoid biased variance estimates.

- Jackknife Repeated Replication.
  - This method creates a replicate by dropping a PSU from one stratum and weights up the other PSUs in the stratum to maintain the sampling distribution across the strata.
  - Advantages:
    - More useful for complex estimates than the Taylor Series.
    - Easily applied to user-specified statistics like differences or ratios of domain means.
    - Can handle designs other than paired selection.
    - Accounts for variability due to multiple steps in adjustment more easily than does Taylor Series.
  - Disadvantages:
    - Not appropriate for the variance of quantiles like the median.
    - Appending replicate weights to each record increases file size.
    - Combining of strata and PSUs is sometimes done to reduce number of replicates. This must be done carefully to avoid biased sampling variance estimates.

Glossary

Accuracy The degree of closeness an estimate has to the true value.

Adjudication The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/adjudication is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to client," "ready to go to fielding agency," for example).

Adjudicator The person who signs-off on a finalized version of a questionnaire (see Adjudication).

Anonymization Stripping all information from a survey data file that allows to re-identify respondents (see confidentiality).

Attitudinal question A question asking about respondents' opinions, judgments, emotions, and perceptions. These cannot be measured by other means; we are dependent on respondents' answers. Example: *Do you think smoking cigarettes is bad for the smoker's health?*

Audit trail An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.

Auxiliary data Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in imputation techniques.

Base weight The inverse of the probability of selection.

Bias The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.

Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: *Do you smoke?* *Yes* ___ *No* ___
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.
Codebook	A document that provides question-level metadata that is matched to variables in a dataset. Metadata include the elements of a data dictionary, as well as basic study documentation, question text, universe statements (the characteristics of respondents who were asked the question), the number of respondents who answered the question, and response frequencies or statistics.
Code structure	List of descriptions of variable categories and associated code numbers.
Coding	Translating nonnumeric data into numeric fields.
Cohen's kappa	A statistical measure that accounts for degree of chance of agreements between coders.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values.
Complex survey data (or designs)	Survey datasets (or designs) based on stratified single or multistage samples with survey weights designed to compensate for unequal probabilities of selection or nonresponse.
Computer assisted personal interviewing (CAPI)	A face-to-face interviewing mode in which a computer displays the questions onscreen, the interviewer reads them to the respondent, and enters the respondent's answers directly into the computer.

Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Contact attempt record	A written record of the time and outcome of each contact attempt to a sample unit.
Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two of more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Cooperation rate	The proportion of all elements interviewed of all eligible units ever contacted.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
Coverage bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value because some elements in the target population do not appear on the sampling frame.
Coversheet	Electronic or printed materials associated with each element that identify information about the element, e.g., the sample address, the unique identification number associated with an element, and the interviewer to whom an element is assigned. The coversheet often also contains an introduction to the study, instructions on how to screen sample members and randomly select the respondent, and space to record the date, time, outcome, and notes for every contact attempt.
Data capture	The process of converting data (e.g., from questionnaires, audio/visual recordings, samples, etc.) to an electronic file.

Data dictionary	A document linking the survey instrument (questionnaire) with the dataset, or more abstract question or variable-level metadata including question identifiers (variable names and labels); response category identifiers (value labels), and data types (e.g., F2.0, specifying that the response is a two-digit integer with zero decimal places.
Design effect	The effect of the complex survey design on sampling variance measured as the ratio of the sampling variance under the complex design to the sampling variance computed as a simple random sample of the same sample size.
Disclosure analysis and avoidance	The process of identifying and protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. For any proposed release of tabulations or microdata, the level of risk of disclosure should be evaluated.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., noncontact, refusal, ineligible, complete interview).
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.
Factual question	A question that aims to collect information about things for which there is a correct answer. In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative records. Factual questions can be about a variety of things, such as figure-based facts (date, age, weight), events (pregnancy, marriage), and behaviors (smoking or media consumption). Example: *Do you smoke?*

Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Hadamard matrix	Square arrays of + and – that define balanced half samples. Such matrices exist for any multiple of four. Pluses [+] mean keep the first PSU and minuses [-] keep the second PSU in the stratum. Therefore, the first half sample identified in the matrix below keeps the first PSU in strata 1, 2, 3 and the second PSU in stratum 4.

Hadamard matrix for 4 half samples

Half Sample	Stratum			
	1	2	3	4
1	+	+	+	-
2	+	-	-	-
3	-	-	+	-
4	-	+	-	-

Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Imputation variance	That component of overall variability in survey estimates that can be accounted for by imputation.
Interviewer design effect ($Deff_{int}$)	The extent to which interviewer variance increases the variance of the sample mean of a simple random sample.
Interviewer variance	That component of overall variability in survey estimates that can be accounted for by the interviewers.
Item nonresponse, item missing data	The absence of information on individual data items for a sample element where other data items were successfully obtained.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Measurement error	Survey error (variance and bias) due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.

Metadata Information that describes data. The term encompasses a broad spectrum of information about the survey, from study title to sample design, details such as interviewer briefing notes, contextual data and/or information such as legal regulations, customs, and economic indicators. Note that the term 'data' is used here in a technical definition. Typically metadata are descriptive information and data are the numerical values described.

Microdata Nonaggregated data that concern individual records for sampled units, such as households, respondents, organizations, administrators, schools, classrooms, students, etc. Microdata may come from auxiliary sources (e.g., census or geographical data) as well as surveys. They are contrasted with macrodata, such as variable means and frequencies, gained through the aggregation of microdata.

Mode Method of data collection.

Noncontact Sampling units that were potentially eligible but could not be reached.

Non-interview A sample element is selected, but an interview does not take place (for example, due to noncontact, refusal, or ineligibility).

Nonresponse The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.

Nonresponse bias The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.

Open-ended question A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers. Example: *What is your occupation?*
Please write in the name or title of your occupation___________

Outcome rate A rate calculated based on the study's defined final disposition codes that reflect the outcome of specific contact attempts before the unit was finalized. Examples include response rates (the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.), cooperation rates (the proportion of all units interviewed of all eligible units ever contacted), refusal rates (the proportion of all units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units), and contact rates (the proportion of all units are reached by the survey).

Outlier An atypical observation which does not appear to follow the distribution of the rest of a dataset.

Overediting Extensive editing that becomes too costly for the amount of error that is being reduced.

Paradata Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.

Poststratification A statistical adjustment that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection.

Post-survey adjustments Adjustments to reduce the impact of error on estimates.

Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less precision is reflected by a larger variance.
Precoding	When designing the questionnaire and survey instrument, determine coding conventions and formats of survey items (especially the closed-ended questions) based on existing coding frames or prior knowledge of the survey population.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Probability proportional to size (PPS)	A sampling method that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for postratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection.
Probability sampling	A sampling method where each element on the sampling frame has a known, non-zero chance of selection.
Processing error	Survey error (variance and bias) that arise during the steps between collecting information from the respondent and having the value used in estimation. Processing errors include all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in the implementation of systems and estimates. In survey data, processing errors may include errors of transcription, errors of coding, errors of data entry, errors in the assignment of weights, errors in disclosure avoidance, and errors of arithmetic in tabulation.
Public use data files	An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze.

Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Recontact	To have someone other than the interviewer (often a supervisor) attempt to speak with the sample member after a screener or interview is conducted, in order to verify that it was completed according to the specified protocol.
Refusal rate	The proportion of all units of all potentially eligible sampling units in which a respondent sampling unit refuses to do an interview or breaks off interviews of all potentially eligible sampling units.
Reliability	The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.
Replicates	Systematic probability subsamples of the full sample.

Response options	The category, wording, and order of options given with the survey question.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.
Sample design	Information on the target and final sample sizes, strata definitions and the sample selection methodology.
Sample element	A selected unit of the target population that may be eligible or ineligible.
Sampling bias	The systematic difference between the expected value (over all conceptual trials) of an unweighted sample estimate and the target population value because some elements on the sampling frame have a higher chance of selection than other elements.
Sampling error	Survey error (variance and bias) due to observing a sample of the population rather than the entire population.
Sampling error computational units (SECUs)	PSUs in 'one PSU per stratum' sampling designs that are grouped in pairs, after data collection, for purposes of estimating approximate sampling variances.
Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.

Sampling variance	A measure of how much a statistic varies around its mean (over all conceptual trials) as a result of the sample design only. This measure does not account for other sources of variable error such as coverage and nonresponse.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Simple random sampling (SRS)	A procedure where a sample of size n is drawn from a population of size N in such a way that every possible sample of size n has the same probability of being selected.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Stratification	A sampling procedure that divides the sampling frame into mutually exclusive and exhaustive groups (or strata) and places each element on the frame into one of the groups. Independent selections are then made from each stratum, one by one, to ensure representation of each subgroup on the frame in the sample.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Survey weight	A statistical adjustment created to compensate for complex survey designs with features including, but not limited to, unequal likelihoods of selection, differences in response rates across key subgroups, and deviations from distributions on critical variables found in the target population from external sources, such as a national Census.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample data.

Taylor Series variance estimation	A commonly used tool in statistics for handling the variance estimation of statistics that are not simple additions of sample values, such as odds ratios. Taylor series handles this by converting a ratio into an approximation that is a function of the sums of the values.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.
Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.
Unique Identification Number	A unique number that identifies an element (e.g. serial number). That number sticks to the element through the whole survey lifecycle and is published with the public data set. It does not contain any information about the respondents or their addresses.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
Universe statement	A description of the group of respondents to which the survey item applies (e.g., "Female, ≥ 45, Now Working").
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse.

References

[1] American Association for Public Opinion Research (AAPOR). (2009). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (6th ed.). Retrieved April 2, 2010, from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814

[2] Bethlehem, J. G. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. M. Groves, A. D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (ch. 18). New York, NY: John Wiley & Sons.

[3] 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.* New York/London: Lawrence Erlbaum Associates.

[4] Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: John Wiley & Sons.

[5] Brogan, D. (2005). Chapter XXI:Sampling error estimation for survey data. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), *Household surveys in developing and transition countries*. New York, NY: United Nations. Retrieved April 15, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[6] Dow, M. M., & Eff, E. A. (2009). Multiple imputation of missing data in cross-cultural samples. *Cross-Cultural Research, 43*(3), 206-229. Retrieved April 15, 2010, from http://ccr.sagepub.com/cgi/reprint/43/3/206

[7] Eff, E. A., & Dow, M. M. (2009). How to deal with missing data and Galton's problem in cross-cultural survey research: A primer for R. *Structure and Dynamics, 3*(3). Retrieved January 10, 2010, from http://www.escholarship.org/uc/item/7cm1f10b

[8] Elias, P. (1997). Occupational classification (ISCO-88): Concepts, methods, reliability, validity and cross-national comparability. *Labour Market and Social Policy Occasional Papers* (No. 20). Paris: OECD.

[9] European Commission. (2010). *Report on Greek government deficit and debt statistics. Brussels: European Commission.* Retrieved February 7, 2010, from http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/COM_2010_REPORT_GREEK/EN/COM_2010_REPORT_GREEK-EN.PDF

[10] Federal Committee on Statistical Methodology. (1983). *Statistical policy working paper 9: Contracting for surveys*. Washington, DC: Office of Management and Budget.

[11] Fellegi, I. P., & Holt, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association, 71*(353), 17-35.

[12] Gabler, S., & Häder, S. (1997): Deviations from the population and optimal weights. In W. E. Saris, & M. Kaase (Eds.) *Eurobarometer: measurement instruments for opinions in Europe* (ZUMA-Nachrichten Spezial 2) (pp. 32-44). Mannheim, Germany: ZUMA. Retrieved February 7, 2010, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial2.pdf

[13] Granquist, L. & Kovar, J. (1997). Editing of survey data: How much is enough? In L. E. Lyberg, P. P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz. (Eds.), *Survey measurement and process quality* (pp. 415-435). New York, NY: John Wiley & Sons.

[14] Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[15] International Statistical Institute (ISI). (2009). *International Statistical Institute declaration on professional ethics*. The Hague, The Netherlands: ISI. Retrieved March 01, 2010, from http://isi.cbs.nl/ethics0index.htm

[16] Jordan, K., Porcheret, M., & Croft, P. (2004). Quality of morbidity coding in general practice computerized medical records: A systematic review. *Family Practice, 21*(4), 396-412.

[17] Kalton, G. (1983). *Compensating for missing survey data*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

[18] Kalton, G., & Kasprzyk, D. (1986). Treatment of missing survey data. *Survey Methodology, 12*, 1-16.

[19] Kessler, R. C., Ustün, T. B., & World Health Organization. (2008). *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge; New York :Geneva: Cambridge University Press; Published in collaboration with the World Health Organization.

[20] Kish, L. (1965). *Survey sampling.* New York, NY: John Wiley & Sons.

[21] Lepkowski, J. M., & Bowles, J. (1996). Sampling error software for personal computers. *The Survey Statistician*, *35*, 10-17.

[22] Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley & Sons.

[23] Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

[24] Marker, D. A, Judkins, D. R., & Winglee, M. (2002). Large-scale imputation for complex surveys. In R. M. Groves, D. A. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse*. New York, NY: John Wiley & Sons.

[25] Office of Management and Budget. (2006). *Standards and guidelines for statistical surveys*. Washington, DC: Office of Information and Regulatory Affairs, OMB. Retrieved January 10, 2010, from http://www.whitehouse.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf

[26] Rawski, T. G. (2001). What is happening to China's GDP statistics? *China Economic Review. 12*(4), 347-354.

[27] Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

[28] Särndal, C. E. and Lundström, S. (2005). *Estimation in surveys with nonresponse*. *New* York, NY: John Wiley & Sons.

[29] Statistics Canada. (2009). *Statistics Canada quality guidelines.* Ottawa: Statistics Canada. Retrieved January 10, 2010, from http://www.statcan.ca/english/freepub/12-539-XIE/index.htm

[30] Singh, M. P., Hidiroglou, M. A., Gambino, J. G., & Kovačević, M. S. (2001). Estimation methods and related systems at Statistics Canada. *International Statistical Review / Revue Internationale de Statistique, 69*(3), 461-486.

[31] United Nations. (2005). *Household surveys in developing and transition countries.* New York, NY: United Nations, Department of Economic and Social Affairs. Retrieved March 26, 2010, from http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf

[32] van de Vijver, F.J.R., & Chasiotis, A. (2010). Making methods meet: Mixed designs in cross-cultural research. In J.A. Harkness, M. Braun, B. Edwards, T.P. Johnson, L. Lyberg, P. Mohler, B-E. Pennell & T.W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 455-473). Hoboken, NJ: John Wiley & Sons.

[33] Vermaak, C. (2009). *The impact of multiple imputation on estimates of the working poor in South Africa.* Paper presented at the Development Policy Research Unit Conference "The Regulatory Environment and Its Impact on the Nature and Level of Economic Growth in South Africa."

[34] Wurdeman, K. (1993). *Quality of data keying for major operations of the 1990 census.* Paper presented at the Proceedings of American Statistical Association: Survey Research Methods Section. Retrieved April 15 2010, from http://www.amstat.org/Sections/Srms/Proceedings/papers/1993_049.pdf

Internet Links

[35] Afrobarometer Survey. Retrieved July 1, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[36] Asian Barometer. Retrieved July 1, 2010, from http://www.asianbarometer.org/newenglish/introduction/default.htm

[37] Census and Survey Processing System (CSPro). (2010). Retrieved May 6, 2010, from http://www.census.gov/ipc/www/cspro

[38] European Social Survey (ESS). (2010). Retrieved July 10, 2010, from http://www.europeansocialsurvey.org

[39] Imputation and Variance Estimation Software (IVEWare). (2009). Retrieved May 6, 2010, from http://www.isr.umich.edu/src/smp/ive/

Further Reading

Groves, R. M. (1989). *Survey errors and survey costs*. Hoboken, NJ: John Wiley & Sons.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.), (2002). *Survey nonresponse*. New York, NY: John Wiley & Sons.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663-665.

Kish, L., & Hess, I. (1959). On variances of ratios and their differences in multistage samples. *Journal of the American Statistical Association*, *54*(286), 416-446.

Lessler, J., & Kalsbeek, W. (1992). *Nonsampling error in surveys*. New York, NY: John Wiley & Sons.

Lyberg, L. E., Biemer, P. P., Collins, M., de Leeuw, E. D., Dippo, C., Schwarz, N. et al. (Eds.). (1997). *Survey measurement and process quality*. New York, NY: John Wiley & Sons.

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27*(1), 85-95.

Smith, T. W. (2003). A review of methods to estimate the status of cases with unknown eligibility. *Report of the Standard Definitions Committee for the American Association for Public Opinion Research.*

Worcester, R., Lagos, M., & Basañez, M. (2000). *Problems and progress in cross-national studies: Lessons learned the hard way.* Paper presented at the 2000 AAPOR/WAPOR Conference, Portland, OR.

XV. Data Dissemination

Peter Granda and Emily Blasczyk

Introduction

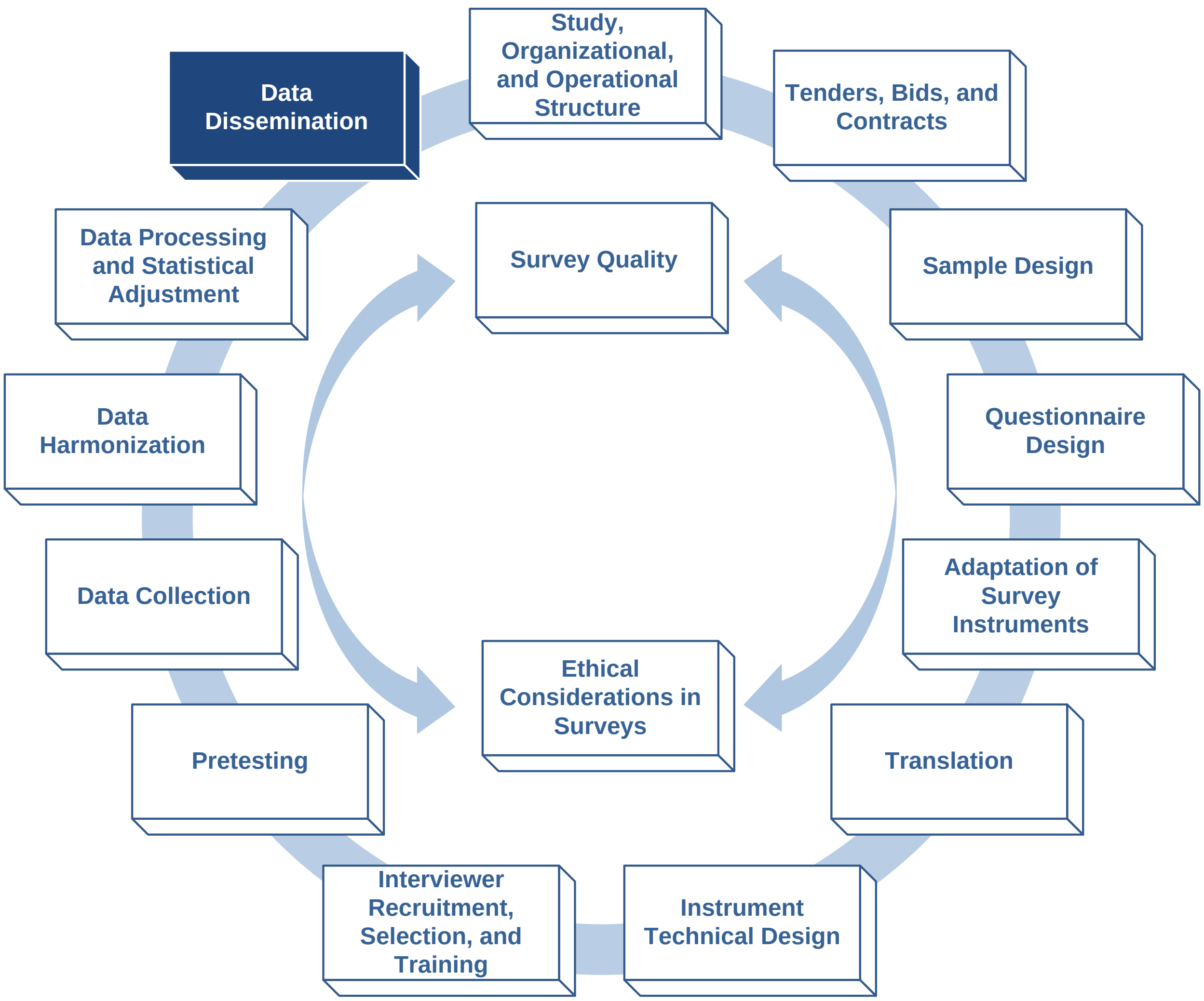
Dissemination is the process by which producers of microdata from surveys and from public and official statistics make their data available to other users. These users may include government officials, academic researchers, policymakers, and the general public. Data may be disseminated publicly without any restrictions (public use files) or only to certain users under specific conditions. The availability of microdata is often dependent on national laws and regulations. Data and documentation may be disseminated in various formats but the goal that producers should have is that the information they provide to others is complete, that the format is not proprietary, and that it is amenable to long-term preservation.

Dissemination of survey data requires careful consideration of several aspects of the process of making data and documentation files available to analysts. More is involved in the dissemination process than merely providing data access to interested researchers. Data producers and archives must assure analysts that the data they provide accurately reflects the efforts of the data collection process, is trustworthy, fully documented, has no confidentiality concerns, and is securely preserved for future use. Disseminating cross-cultural survey data includes specific processes such as standardization, harmonization, and multi-lingual documentation which may not apply to surveys done in a single country.

An additional aspect of dissemination is how to share research findings with interested parties. Determining who is using the data and why they are using it is important to consider as part of a comprehensive dissemination strategy. Many international organizations, social science data archives, and survey research projects also embrace these objectives. Although focused on micro-economic data, the International Monetary Fund, for example, established a set of guidelines on macroeconomic data for member countries to follow in order to provide the public with "comprehensive, timely, accessible, and reliable economic, financial, and socio-demographic data" [8] [5].

Figure 1 shows data dissemination within the survey production process lifecycle (survey lifecycle) as represented in these guidelines. The lifecycle begins with establishing study structure (Study, Organizational, and Operational Structure) and ends with data dissemination. In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle. Survey quality can be assessed in terms of fitness for intended use (also known as fitness for purpose), total survey error, and the monitoring of survey production process quality, which may be affected by survey infrastructure, costs, respondent and interviewer burden, and study design specifications (see Survey Quality).

Figure 1. The Survey Lifecycle



Guidelines

Goal: To ensure that survey and statistical research teams in all cultures and countries involved in a project follow accepted standards for the long-term preservation and dissemination of data to the social science research community and the wider public.

1. **Make a dissemination and data preservation plan early in the project lifecycle that includes archiving, publishing, and distribution.**

Rationale

Dissemination is an integral part of modern survey research. It involves the documentation of major steps in the data lifecycle from initial planning to the production of final data files. This includes, when available and appropriate, detailed information about the survey process (paradata), all data editing steps, and protocols which determine what types of data and documentation files are made available to which users.

Procedural steps

- For cross-cultural surveys, decide on the standard documentation language to be used.

- Identify any documents that should be published in their original language such as individual country questionnaires, codes, and nation-specific data files.

- Have a system in place to preserve all major planning and operational documents as soon as they are created.

- Consider including some information about the survey process when disseminating data, documentation, and reports. Producers may want to balance the amount of paradata they release with the need to maintain proprietary information about the data collection process.

Lessons learned

- All studies must develop a system for preserving and storing materials.
    - Round 4 of the Afrobarometer Survey strongly recommends that participating countries scan their completed paper-and-pencil questionnaires. Hard copies are acceptable where circumstances (e.g., cost) prevent scanning. National partners are responsible for either the scanning or the storing of their own questionnaires. Each national partner is responsible for entering and cleaning their own data and delivering a clean SPSS data set [18].
    - All documents related to Round 5 of the European Social Survey (ESS) are uploaded to a server. This includes, but is not limited to, original unedited (raw) data, fieldwork documents, metadata, and population statistics for coverage and response rates [19].
    - Documentation of International Social Survey Programme (ISSP) survey methods and data files are sent to a central data archive no later than nine months after fieldwork is completed. Data is to be sent unweighted, but descriptions of weighting procedures should accompany the datasets [20].

- Master copies of all important Living Standard Measurement Study Survey (LSMS) files are kept in a separate archive which is backed-up [10].
- Documentation for the World Mental Health (WMH) Survey is done using the Survey Metadata Documentation System designed by the WMH Data Collection Coordination Centre [9].
- Countries participating in the World Value Survey are required to submit documentation of their survey methods and data to a central data archive no later than three months after fieldwork has been completed. Documentation must include a completed methodology questionnaire, a report of any questions omitted or added to the original official questionnaire, a report of additional and/or country specific codes to any questions, official demographic statistics, weights used, and a copy of the original country questionnaire [17] [22].

- Many institutions which provide research grants for data collection now strongly recommend that grantees prepare a data sharing plan as part of the proposal process. The National Institutes of Health in the United States (NIH) provide the following justification for their emphasis on dissemination: “Data sharing promotes many goals of the NIH research endeavor. It is particularly important for unique data that cannot be readily replicated. Data sharing allows scientists to expedite the translation of research results into knowledge, products, and procedures to improve human health. There are many reasons to share data from NIH-supported studies. Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new datasets when data from multiple sources are combined.” This policy has resulted in more data becoming available in the public domain.

2. **Preserve sustainable copies of all key data and documentation files produced during the data collection process, as well as those made available for secondary analyses.**

Rationale

Preservation is an important part of the survey lifecycle, a prerequisite for long-term access to valuable physical objects and digital materials. The materials that need to be preserved and kept available to members of the research community include such objects as public use data and documentation files (including key files used in their construction), copies of the data collection instruments, user guides, information about the data collection process, and reports on field operations. Since dissemination policies may differ among countries, it is important that data producers take the necessary steps to make their collections as accessible as possible to members of the research community. This may include organizing dissemination themselves.

Procedural steps

- Define the long-term preservation standards and protocols to be used. Consider digitizing physical objects, commonly-used questionnaires, or other administrative materials documenting the whole data lifecycle including the design phase.

- Observe the "date of expiry" of storage and data formats.

- Identify storage devices that are certified for long-term data preservation. Clearly mark the "recopy by …" date.

- Protect digital materials through storage of multiple copies in multiple locations. An ideal preservation storage situation includes a minimum of several off-site copies of digital materials undergoing regularly scheduled back-ups. If it is not possible to store materials at multiple sites, preserve at least one copy in a different location.

- Make certain that digital materials remain retrievable through constant refreshment of the media on which they are stored. This is particularly important if removable media such as tapes are used for storage, since formats and the machines required to read these media change quickly over time.

- Maintain older versions of important data and documentation files so users can follow the changes made from one version to the next.

- At a minimum, store a copy of all data and metadata files in software-independent formats such as ASCII files and XML which, with proper accompanying documentation, can be read into all major statistical packages.

- Investigate the protocols and standards of digital repositories, such as availability of extracting data and in the areas of multi-site storage, security, and costs.

Make test runs of copied data to ensure error-free copy processes.

- Work if possible with a trusted digital repository, such as a national or public social science data archive, to preserve all study materials. In doing so, data producers do their best to ensure that their data collections will remain available to the research community.
  - Such repositories make an explicit commitment to preserving digital information by:
    - Complying with the Open Archival Information System (OAIS) in the US and other similar standards in other countries which have their own digital preservation standards and practices [11] [13] [16].
    - Ensuring that digital content can be provided to users and exchanged with archives without damaging its integrity.
    - Participating in the development and promulgation of digital preservation community standards, practice, and research-based solutions.
    - Developing a reliable, sustainable, and auditable digital preservation repository that has the flexibility to grow and expand.
    - Managing the hardware, software, and storage media components of the digital preservation function in accordance with environmental standards, quality control specifications, and security requirements

- If no national or public social science data archives exist, consider depositing data with an archive in another country or investigate the possibility of doing so with a national statistical agency or certified provider. Consider archiving collections in one archive which would keep master copies of files in several locations but minimize the possibility of conflicting versions of data and documentation files.

Lessons learned

- Data producers should strongly consider a preservation strategy before putting files online for people to download. For example, many data and documentation files available on Web sites undergo frequent changes and updates. When updates are made, the older version of the files is often no longer available. This may make it difficult, if not impossible, to replicate previous analyses done by the user, or to test the assumptions and results of analyses done by others. A system of version control is necessary to insure that analysts know which files they are using.

- The German National Science Foundation (DFG) requires data to be archived for a minimum of 10 years as part of its anti-fraud activities.

- Some earlier studies did not preserve individual country data, thus issues about harmonization emerging some decades later could not be settled.

- Data producers should make every effort to extract data that is on media which may no longer be easy to read. Too many data files have been irretrievably lost because no one bothered to copy them to newer types of media.

3. **Conduct effective disclosure analysis to protect respondent confidentiality.**

Rationale

Any plan to disseminate survey data must include very specific procedures for understanding and minimizing the risk of breaching the promise of confidentiality that is made to respondents at the time of the survey or collection of data. The key goal of disclosure risk analysis and processing is to ensure that the data maintain the greatest potential usefulness while simultaneously offering the strongest possible protection to the confidentiality of the individual respondents.

Procedural steps

- Implement a disclosure protocol. A proper disclosure protocol includes an analysis of the most likely outside sources which might allow the identification of respondents or households.

- Search systematically in the data file for sensitive information such as transcripts of open ended answers including ISCO (International Standard Classification of Occupations) occupational variables, identification of PSUs, birth dates, income, or housing and dwelling information.

- Search also for unusual characteristics and for cells in tables with very low frequencies.

- Undertake both practical and statistical steps to identify cases and variables. This allows the identification of areas or variables that need to be further masked in order to prevent identification of subjects, either through analysis or by matching study data with data from other external databases. After having decided on which variables present unacceptable risks, mask the relevant information.

- Evaluate data files once those cases and variables are identified. In virtually every case, the data can be masked in various ways that make it possible for public use data to be distributed, usually through a Web-based system.

- Use appropriate masking procedures to preserve respondent confidentiality while also trying to optimize the usefulness of the resultant data file for analysis. These procedures might include top or bottom coding of key demographic variables such as income, removing data for very sensitive variables, and swapping data values between similar cases [12].

- Document all confidentiality assurance processes and make a final assessment about the anonymity of the data file.

Lessons learned

- With the enhanced emphasis on privacy in almost all countries, confidentiality reviews of microdata are increasingly important, if not indispensable, to assuring the future availability of public use data.

- Confidentiality is both a theoretical and an empirical issue. There is a tendency to 'over protect' survey data based on theoretical considerations or simulation studies. Careful tests conducted in the 1980s with German Micro Census data showed that previous simulation studies were exaggerating the threat of information disclosure.

- The practice of reporting examples of privacy violations, particularly in the health care field in the United States, has increased awareness of this issue [6].

4. Consider the production of both public- and restricted-use data files.

Rationale

In order to ensure that researchers have access to the greatest amount of data without compromising respondent confidentiality, data producers, when appropriate, must make every effort to create both public- and restricted-data documentation files, and make these files available to the research community through secure and predictable channels.

Procedural steps

- Make data files as fully available to the research community as possible within the confines of how the project is organized and financed. If general distribution is not feasible, establish clear rules under which researchers can obtain the data.

- Remain cognizant of the fact that data files, however they are disseminated, are always 'owned' by the principal investigator(s) who maintain permanent copyright privileges over their products.

- Provide access directly by the data producer if resources permit, but also always send copies to a trusted digital repository for permanent preservation, in case the data producer should cease to provide access at some time in the future.

- Consider the creation of less thoroughly masked versions that can be distributed under restricted-use contracts, or made available within a research data center or "enclave" (i.e., a secure environment in which the user has access to restricted data and analytic outputs under controlled conditions).

- Establish clear policies for how researchers may access restricted data files by creating a set of application materials and restricted-use data agreements that specify how researchers can obtain and use such data [7].

- Distribute restricted files through signed data use agreements. These may incorporate data protection plans, formal licenses, and travel to a special facility at which researchers can access the data in a very controlled environment.

- Create special files for researchers that cannot be matched with public use files (for example, provide finer grained local information and simultaneously change respondents' IDs and other matching variables).

- In order to provide optimal utility for researchers, produce a variety of products for varied constituencies.
  - Produce setup files and ready-to-use 'portable' files in SAS, SPSS, and Stata to address the needs of those who seek to do intensive statistical analyses with particular software packages.
  - Consider disseminating data on removable media, e.g., CD-ROM or DVD if appropriate.

- Address the needs of policymakers and those who are browsing for new data sources, seeking summary analytic information, or wanting to quickly download specific variables by creating tools within the Web-based system to permit online analysis, subsetting, and access to full documentation. Be aware that online analysis must use fully anonymized data.

Lessons learned

- Established cross-cultural studies share their data in a variety of ways. The Afrobarometer Survey releases all data, along with relevant codebooks, via website and other outlets; to allow initial in-house analysis and publication, data is not released publicly until one year after the completion of fieldwork [18]. Anonymized data are released onto the European Social Survey (ESS) public website within one year of the onset of data collection [19]. The International Social Survey Programme (ISSP) makes individual national and/or combined datasets available to the scientific community by the Data Archive one year after the calendar year to which it relates [20]. Living Standard Measurement Study Survey (LSMS) data is usually available within six to eighteen months of the end of fieldwork and is published in the LSMS Working Paper series [10]. Survey of Health, Ageing, and Retirement in Europe (SHARE) data is distributed through their Research Data Center [21]. Collected data from the World Value Survey is available only to participating countries for a period of two years after fieldwork has been completed; after this period, the data is made available to the worldwide social science community in the form of data archives [17] [22].

- Consider making clear agreements on data heritage (i.e., copyright transfer after the original principal investigator retires). A German elite study was nearly lost to the academic public due to heritage issues.

- Most data are already paid for by taxpayer money or foundations. Thus foundations and public funders often ask for free data access (i.e., they deny the principal investigator's sole ownership on collected data).

- Despite general agreement about the advantages of making data accessible to other researchers, as well as strong data-sharing cultures in many nations, too few social science data collections are effectively preserved. Data archives should do as much as possible to facilitate the deposit process by contacting principal investigators and data producers as they prepare data and documentation files.

- More than ten years ago the International Monetary Fund (IMF) began to develop a set of dissemination standards "to guide countries in the provision to the public of comprehensive, timely, accessible, and reliable economic, financial, and socio-demographic data" [8]. These standards were considered best practices but their implementation was completely voluntary depending on the policies and wishes of each nation. The Fund recently published a report [1] about the success of this initiative over the last decade. It concluded that more accurate and reliable statistical information is now being produced by many nations than ever before but also recognized that dissemination mechanisms are not fully developed in many locations. Nations also have internal challenges and constraints in addressing dissemination goals from resource constraints, shifting priorities, and in their ability to generate periodic and timely statistical data.

5. **Produce data files that are easy for researchers to use.**

Rationale

An effective data processing strategy focuses on the production of data files that will provide optimal utility for researchers. Such files have been thoroughly checked and cleaned, possess uniform and consistent coding strategies, use common formats, and address the potential research needs of secondary analysts.

Procedural steps

Processors should perform a series of steps to ensure the integrity and maximum utility of public-use files. Such steps include:

- Make a thorough investigation of any undocumented code numbers or inconsistent responses. Whenever possible provide labels for such codes such as ‘not ascertained’ if there is no alternative.

- Standardize all missing data values, unless it is not possible to do so because of different cultural understandings (flag such issues carefully). Users doing analyses will appreciate that all “does not apply,” “don’t know,” “refused,” and “no data available” responses are coded the same way in the data file.

- Create complete and concise variable and value labels which will provide researchers with clear descriptions of their analytic results.

- Provide a printable questionnaire that contains all variable names and values in an appropriate format.

- Format the data files in a way that permits access through a wide variety of statistical packages, all of which will produce the same results no matter how complicated the analysis requested, particularly with any variable where decimal precision is an important consideration.

- Consider producing ancillary files for those data collection efforts which cover multiple waves of respondents or several geographic areas. Such files may include recoded variables to summarize information contained in many questions or special constructed variables that producers feel will aid researchers in their analyses.

- Create special subsets of data which take advantage of the longitudinal richness of long-term collections and provide unique opportunities to study important social, political, and economic issues from different perspectives, particularly with regard to the changing characteristics of the sampled respondents.
- Whenever possible and expedient, make individual country datasets available in cross-national surveys.
- Consider creating simplified versions of datasets for use by a wider public such as journalists and policymakers (i.e., by creating recode variables such as age of respondents in groups, income in groups, removing detailed information such as household lists, setting missing data properly, etc.) [15]. Make such datasets accessible via web-analysis.

Lessons learned

- Users increasingly expect data files to come in a variety of formats that will work easily with their statistical package of choice. In some settings this may be just an SPSS portable file, but in others data producers and/or archives might need to create the same file in a variety of formats, particularly if a standard database conversion package, such as STAT-TRANSFER, is not available.
- Be very clear about coding responses that refer to “item response refused,” “item response does not apply due to filtering,” “can’t choose all” or “don’t know,” and especially “no code in data file where a code should be.” All these have different meanings and must get different values. The “no code in data file” indicates either an interviewer error or error in data editing.
- “Don’t know”/”Can’t choose” responses may have different meaning in different countries based on different response styles. Treating all of these responses as missing data may lead to unwarranted conclusions about the attitudes of whole populations [14].

6. **Develop finding aids to guide users in their quest to locate data collections they want to use.**

Rationale

Finding aids are critical to all data dissemination systems, from individual data producers, with only a few data collections, to social science archives with thousands of such collections.

Procedural steps

- Create a robust search engine to query the fielded metadata so that the user can find variables of interest efficiently.

- Allow the search engine to run against a study's bibliography to enable two-way linking between variables and publications based on analyses of those variables.

- Display the abstracts of the publications with links to the full text whenever possible, in order to realize the full potential of the online research environment.

- Dedicate staff time to continuously search journals and online databases to discover new citations where the data have been used.

- Encourage data archives to create metadata records for surveys they do not preserve and distribute these records to facilitate their discovery and use.

Lessons learned

- Use of data increases when the data are easy to find and when users know which publications previous scholars have generated from such data. There are many datasets that would be of interest to secondary analysts if the analysts only knew about them. For example, many surveys were conducted in Latin America and Africa in the 1960s and 1970s which might offer opportunities for interesting comparative analyses with the more recent and much more popular Latino and Afrobarometer surveys. These are not always as visible to researchers, however, as they might not possess immediately obvious substantive or methodological interest.

7. Create comprehensive training, outreach, and user support programs to inform the research community about the dataset.

Rationale

Training and support of users will increase usage of the data and encourage comprehensive analyses. It is very important that major survey research producers or archives reach out to the user community effectively, in order to explain the structure of new datasets and to encourage the greatest possible use. The most straightforward way to reach out is to develop an effective online presence, ensuring that the data are easily located and acquired, and that metadata and bibliographical citations are also available. Good user support will prevent obvious misuse or possible misunderstanding of the structure and content of the dataset.

Procedural steps

- Organize workshops at relevant professional organizations or plan conferences soon after the data are released, in order to bring early users together to discuss important preliminary results, as well as to ensure that the data are used effectively and that any problems with the data are recognized and corrected.

- Hold training workshops to ensure that novice users have a chance to learn about the data from experts and, if possible, from the data production team itself.
  - Without specialized instruction and training, analyses of cross-cultural, longitudinal data and repeated cross-sectional data are particularly challenging.
  - These training courses can be brief half-day or one-day sessions at the time of professional meetings, or they can continue for longer periods (e.g., three- or five-day sessions with a more detailed focus).

- Send representatives to important professional meetings with a display "booth," where staff from the project can describe the data, distribute documentation and sample data, and encourage researchers to make use of the data.

- Provide easy access to user support through phone, email, online chat, user forums, and tutorials.

- Track all user questions in a database that creates an accumulating knowledge base and that can also serve to generate Frequently Asked Questions.

- Create tutorials, some of which may be offered in video format, to provide help in using the data, the online analysis system, and the major statistical software packages.

- Establish moderated user forums to provide the foundation for an online community of researchers and students who can discuss their experiences using data and learn from each other.

Lessons learned

- Training programs must be well-planned, with a high level of substantive, methodological, and technical expertise, in order for participants to benefit from the experience. While data producers are usually those who best understand their data, they may not have the resources or desire to provide ongoing user support for the research community. Some may delegate this task to a data archive, but a joint approach, with data archives providing basic user support and data producers addressing more complicated substantive questions, often works best.

- Complex data sets often require specialized training. Data collection methods or sampling frames often change between different waves or in different countries and weighting variables may require extensive descriptions. In this context, there is no real substitute for intensive training and ongoing user support.

8. Produce comprehensive documentation for all public use data files.

Rationale

High-quality documentation is essential for effective data use. Data producers must strive to provide documentation, commonly referred to as metadata, on all aspects of the survey or statistical life cycle, from initial planning through final data production and its release to the research community. (For more information on data processing techniques, see Data Processing and Statistical Adjustment.)

Procedural steps

- Keep good records from the very beginning of the project and make every attempt to record important project events at the time they occurred. This will assist analysts in understanding the goals and purpose of each survey.

- Update documentation continually during the entire lifecycle of the project and preserve old versions of key files.

- For cross-national surveys, provide complete information about how the survey was conducted in each country, and describe specific procedures and practices involving data collection and data processing activities.

- Consider adopting the Data Documentation Initiative [3] standard for producing metadata. The use of this emerging standard, which is based on the use of XML (eXtensible Markup Language), allows specification of each metadata element (e.g., title of the survey, name of the principal investigators, type of sampling) for storage and future searching.

- Define a database structure that will be used to store XML elements.

- Identify appropriate tools that will access and create XML coded information in a natural language environment such as a browser displaying a web-based form generator.

Lessons learned

- XML metadata markup offers new opportunities for data producers to create their documentation, as well as several advantages to users of the documentation:
  - All information that the analyst needs is available in a core document, from which other products, such as text files that contain the necessary information to run statistical analyses in software programs, can be produced.
  - The XML file can be viewed with Web browsers and lends itself to Web display and navigation.
  - Because the content of each field of the documentation is tagged, the documentation can serve as the foundation for extract and analysis programs, search engines, and other software agents written to assist the research process.
  - Preparing documentation in DDI format at the outset of a project means that the documentation will also be suitable for archival deposit and preservation, because it will contain all of the information necessary to describe all of aspects of the corresponding data files. DDI XML should ideally be generated by the CAI system used to collect data, but can also be collected from paper and pencil surveys through access to the information in the original questionnaire.
  - Although few principal investigators of survey data have yet produced full DDI-compliant metadata, the few examples [3] that exist illustrate the importance of using this developing standard at the variable level. New use cases, currently under preparation, will demonstrate additional features, such as:
    - The presentation of instrument documentation, so that users can track the logic of the questionnaire.

- The creation of questions banks, comprising everything asked in multi-year studies, years they were asked, differences in question wording, and so on. XML marked up information gets its full potential when coupled with a database management system and powerful front end tools.
- The establishment of links to the documentation of related surveys (e.g., those conducted in other countries) with variable text viewable in the native languages, assists analysts who want to study relationships among all of the survey items.

9. Consider disseminating research findings.

Rationale

Dissemination is more than storing (archiving) data. Presenting research findings in addition to making the data file available to other users is an important step in quality dissemination practices. This section of the chapter discusses dissemination in terms of presenting results of the study, and considering who will use the information and why. This guideline is based on the guidelines written by the Community Advisory Board of the University of California, San Francisco, Center for AIDS Prevention Studies (CAB CAPS) and is adapted for the cross-cultural context [4].

Procedural Steps

- Create a dissemination plan
  - Include presenting findings in the study's budget. This may include salary, translation, printing, mailing, and/or meeting costs (see Tenders, Bids, and Contracts and Translation).
  - Create a team which will organize and create materials.
  - Get input from study participants, community representatives, and other potentially interested parties on the preferred forum for viewing findings, such as press releases, websites, newsletters, or conferences. Consider offering multiple venues, if possible.
  - Remember that there may be a need to disseminate findings several times as new information is collected and updated.

- Make research results accessible to the desired audience(s). Potential audiences and effective methods include:
  - Study participants:
    - Ask participants if and how they would want to receive results. This can be incorporated as a question in the survey instrument.
    - Create a newsletter for participants.

    - Write any information disseminated in accessible language, and keep in mind the literacy and language needs of the study population.
  - Community members/Target populations:
    - Consider multiple methods including articles in the media such as newspapers, radio, or TV news in order to reach many people.
    - As with study participants, consider the language needs of the community.
    - Explore how research results from cross-national surveys can be disseminated to as many participating countries as possible. Different dissemination strategies may need to be employed in different countries/cultures.
  - Agencies and Service Providers:
    - Prioritize contacting agencies that aided with participant recruitment and/or serve the target population.
    - Emphasize practical use of the study results.
  - Policymakers:
    - Evaluate if research results have potential to impact policy.
    - Send newsletters/articles or reports to local and national government representatives.

- Consider the ethical and legal policies within each country and culture. Individual countries may have different dictates on sharing data within and between countries. (See Ethical Considerations.)

Lessons Learned

- Traditionally, researchers disseminate work in peer-reviewed journals. However, practitioners, as well as the general public, rarely have the time, or even ability, to read these types of articles. The CAB CAPS guidelines were created by a committee of activists, teachers, and other stakeholders. Committee members who had participated in research studies were concerned over the lack of accessible findings, and developed the above points in order to address dissemination needs. Making the attempt to disseminate work in this way provides more benefit to those who funded the research project and encourages discussion about the strengths and weaknesses of the original data.

- The Afrobarometer Survey issues first reports or bulletins within three months of the end of fieldwork. An advance briefing is offered to top policy makers in the executive and legislative branches of participating countries; immediately thereafter, results are released publicly to the national and international media, civil society, and donors. Releases must be approved by a core partner [18]. Similarly, data from the

World Mental Health Survey is available to policy makers in participating countries [9].

10. Make quality control an integral part of all dissemination steps.

Rationale

Dissemination requires strict compliance to archiving, editing, publishing, and distribution protocols. Dissemination also requires the long-term availability of data and documentation files though constantly new versions of hardware, software, and possible changes in management and staff. Clear procedures must be in place to make certain all files are readable as statistical and word processing software systems change over time.

Procedural steps

- Establish a quality compliance protocol.
- Check all dissemination production steps throughout.
- Test archived files periodically to verify user accessibility.
- Establish procedures early in the survey lifecycle to insure that all important files are preserved.
- Create digitized versions of all project materials, whenever feasible.
- Develop specific procedures for assessing disclosure risk to respondents and execute these procedures whenever public-use files are produced.
- Produce and implement procedures to distribute restricted-use files if applicable.
- Provide data files in all the major statistical software packages and test all thoroughly before they are made available for dissemination.
- Designate resources to provide user support and training for secondary researchers.
- Discuss with users their experiences working with the data. This may include surveying users, conference presentations, and collecting user data.

Lessons learned

- The Centers for Disease Control (CDC) in the United States, working with other federal agencies, did a study of web-based systems for the dissemination of health data and produced a *Guide for Public Health Agencies Developing, Adopting, or Purchasing Interactive Web-based Data Dissemination Systems.* The *Guide* was developed based on the experiences of many health agencies in disseminating their data and attempts to establish a set of general standards and practices. A checklist is provided to guide agencies in developing a comprehensive web dissemination system [2].

Glossary

Anonymity	Recording or storing information without name or identifier, so the respondent cannot be identified in any way by anyone. No one can link an individual person to the responses of that person, including the investigator or the interviewer. Face-to-face interviews are never anonymous since the interviewer knows the address (and likely, the name) of the respondent.
Anonymization	Stripping all information from a survey data file that allows the re-identification of respondents (see confidentiality).
ASCII files	Data files in American Standard Code for Information Interchange (ASCII) format.
Audit trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey (timestamps) and interviewer or respondent actions while proceeding through a survey. The file may contain a record of keystrokes and function keys pressed, as well as mouse actions.
Auxiliary data	Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in imputation techniques.
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Bottom coding	A type of coding in which values that exceed the predetermined minimum value are reassigned to that minimum value or are recoded as missing data.
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster.

Coding	Translating nonnumeric data into numeric fields.
Complex survey data (or designs)	Survey datasets (or designs) based on stratified single or multistage samples with survey weights designed to compensate for unequal probabilities of selection or nonresponse.
Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Constructed variable	A recoded variable, one created by data producers or archives based on the data originally collected. Examples are age grouped into cohorts, income grouped into 7 categories, Goldthorpe-Index, or the creation of a variable called POVERTY from information collected on the income of respondents.
Coverage	The proportion of the target population that is accounted for on the sampling frame.
Data Documentation Initiative (DDI)	An international effort to establish a standard for technical documentation describing social science data. A membership-based Alliance is developing the DDI specification, which is written in XML.
Disclosure analysis and avoidance	The process of identifying and protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. For any proposed release of tabulations or microdata, the level of risk of disclosure should be evaluated.
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking consistency, correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes coding and imputation, the placement of a number into a field where data were missing.

Fitness for intended use	The degree to which products conform to essential requirements and meet the needs of users for which they are intended. In literature on quality, this is also known as "fitness for use" and "fitness for purpose."
Imputation	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Inconsistent responses	Inappropriate responses to branched questions. For instance, one question might ask if the respondent attended church last week; a response of "no" should skip the questions about church attendance and code the answers to those questions as "inapplicable." If those questions were coded any other way than "inapplicable," this would be inconsistent with the skip patterns of the survey instrument.
Item nonresponse, item missing data	The absence of information on individual data items for a sample element where other data items were successfully obtained.
Longitudinal study	A study where elements are repeatedly measured over time.
Mean Square Error (MSE)	The total error of a survey estimate; specifically, the sum of the variance and the bias squared.
Metadata	Information that describes data. The term encompasses a broad spectrum of information about the survey, from study title to sample design, details such as interviewer briefing notes, contextual data and/or information such as legal regulations, customs, and economic indicators. Note that the term ‘data’ is used here as a technical definition. Typically metadata are descriptive information and data are the numerical values described.
Microdata	Nonaggregated data that concern individual records for sampled units, such as households, respondents, organizations, administrators, schools, classrooms, students, etc. Microdata may come from auxiliary sources (e.g., census or geographical data) as well as surveys. They are contrasted with macrodata, such as variable means and frequencies, gained through the aggregation of microdata.

Nonresponse	The failure to obtain measurement on sampled units or items. See unit nonresponse and item nonresponse.
Paradata	Empirical measurements about the process of creating survey data themselves. They consist of visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, and observations of respondents themselves about the data collection. Examples include timestamps, keystrokes, and interviewer observations about individual contact attempts.
'Portable' file	A file that is coded in a non-proprietary format such as XML or ASCII and thus can be used by a variety of software and hardware platforms.
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Public use data files	An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the survey lifecycle are of the highest achievable quality. Quality assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the quality management plan has clearly outlined quality assurance, quality control, corrective actions to be taken, etc., and whether they have been effectively carried out.

Quality control

A planned system of process monitoring, verification and analysis of indicators of quality, and updates to quality assurance procedures, to ensure that quality assurance works.

Quality management plan

A document that describes the quality system an organization will use, including quality assurance and quality control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.

Response rate

The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.

Response styles

Consistent and stable tendencies in response behavior which are not explainable by question content or presentation. These are considered to be a source of biased reporting.

Restricted-use data files

A file that includes information that can be related to specific individuals and is confidential and/or protected by law. Restricted-use data files are not required to include variables that have undergone coarsening disclosure risk edits. These files are available to researchers under controlled conditions.

Sample design

ormation on the target and final sample sizes, strata definitions and the sample selection methodology.

Sample element

A selected unit of the target population that may be eligible or ineligible.

Sampling frame

A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.

Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the sampling units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term primary sampling units (PSUs) refers to the sampling units chosen in the first stage of selection. The term secondary sampling units (SSUs) refers to sampling units within the PSUs that are chosen in the second stage of selection.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Strata (stratum)	Mutually exclusive, homogenous groupings of population elements or clusters of elements that comprise all of the elements on the sampling frame. The groupings are formed prior to selection of the sample.
Stratification	A sampling procedure that divides the sampling frame into mutually exclusive and exhaustive groups (or strata) and places each element on the frame into one of the groups. Independent selections are then made from each stratum, one by one, to ensure representation of each subgroup on the frame in the sample.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Survey weight	A statistical adjustment created to compensate for complex survey designs with features including, but not limited to, unequal likelihoods of selection, differences in response rates across key subgroups, and deviations from distributions on critical variables found in the target population from external sources, such as a national Census.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.

Timestamps — Timestamps are time and date data recorded with survey data, indicating dates and times of responses, at the question level and questionnaire section level. They also appear in audit trails, recording times questions are asked, responses recorded, and so on.

Top coding — A type of coding in which values that exceed the predetermined maximum value are reassigned to that maximal value or are recoded as item-missing data.

Total Survey Error (TSE) — Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the estimation and reduction of the mean square error (MSE) of statistics of interest.

Trusted digital repository — A repository whose mission is to provide reliable, long-term access to managed digital resources to its designated community, both now and in the future.

Undocumented code number — A code that is not authorized for a particular question. For instance, if a question that records the sex of the respondent has documented codes of "1" for female and "2" for male and "9" for "missing data," a code of "3" would be an "undocumented code."

Unit nonresponse — An eligible sampling unit that has little or no information because the unit did not participate in the survey.

Variance — A measure of how much a statistic varies around its mean over all conceptual trials.

Weighting — A post-survey adjustment that may account for differential coverage, sampling, and/or nonresponse processes.

XML (eXtensible Markup Language) — XML (Extensible Markup Language) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere. XML documents are made up of storage units called entities, which contain either parsed or unparsed data. Parsed data is made up of characters, some of which form character data, and some of which form markup. Markup encodes a description of the document's storage layout and logical structure. XML provides a mechanism to impose constraints on the storage layout and logical structure.

References

[1] Alexander, W. E., Cady, J. & Gonzalez-Garcia, J. (2008). The IMF's data dissemination initiative after 10 years. Washington, DC: International Monetary Fund. Retrieved May 2, 2010, from http://www.imf.org/external/pubs/ft/books/2008/datadiss/dissemination.pdf

[2] Centers for Disease Control and Prevention (CDC). (2010). *Evaluation of State web-based systems for the dissemination of health-related data*. Retrieved January 11, 2010, from http://www.cdc.gov/ncphi/disss/nndss/asb/orcmacro.htm

[3] Data Documentation Initiative (DDI). (2010). Retrieved January 11, 2010, from http://www.ddialliance.org/

[4] Fernández-Peña, J. R., Moore, L., Goldstein, E., DeCarlo, P., Grinstead, O., Hunt, C., et al. (2008). Making sure research is used: Community generated recommendations for disseminating research. Progress in Community Health Partnerships: Research, *Education, and Action, 2*(2), 171-176.

[5] Gutmann, M. P., Abrahamson, M., Adams, M. O., Altman, M., Arms C., Bollen. K. et al. (2009). From preserving the past to preserving the future: The Data-PASS project and the challenges of preserving digital social science data. *Library Trends, 57*(3). Retrieved May 02, 2010, from http://muse.jhu.edu/journals/library_trends/summary/v057/57.3.gutmann.html

[6] Health Insurance Portability and Accountability Act Privacy and Security (HIPAAps). (2010). *Examples of privacy violations*. Retrieved January 11, 2010, from http://www.hipaaps.com/examples.html

[7] Inter-university Consortium for Political and Social Research (ICPSR). (2010). *Data sharing for demographic research.* Retrieved January 11, 2010, from http://www.icpsr.umich.edu/DSDR/rduc/

[8] International Monetary Fund (IMF). (2010). *Dissemination standards bulletin board.* Retrieved January 11, 2010 from http://dsbb.imf.org/Applications/web/dsbbhome/

[9] Kessler, R. C., Ustun, T. B., & World Health Organization. (2008). The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders. Cambridge; New York: Geneva: Cambridge University Press; Published in collaboration with the World Health Organization.

[10] Living Standard Measurement Study Survey. (1996). Working paper 126: A manual for planning and implementing the Living Standard Measurement Study Survey. Retrieved July 10, 2010, from

http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2000/02/24/000009265_3961219093409/Rendered/PDF/multi_page.pdf

[11] National Digital Archive of Datasets (NDAD). (2010). Retrieved January 11, 2010, from http://www.ndad.nationalarchives.gov.uk/

[12] O'Rourke, J. M., Roehrig, S., Heeringa, S. G., Reed, B. G., Birdsall, W.C., Overcashier, M. (2006). Solving problems of disclosure risk while retaining key analytic uses of publicly released microdata. *Journal of Empirical Research on Human Research Ethics, 1*(3), 63-84.

[13] Royal Statistical Society & the UK Data Archive. (2002). *Preserving & sharing statistical material. Colchester, Essex: University of Essex.* Retrieved January 11, 2010, from http://www.rss.org.uk/PDF/Archiving.pdf

[14] Schafer, J.L. & Grahm, J.W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

[15] Sicinski, A. (1970). "Don't Know" answers in cross-national surveys. *Public Opinion Quarterly, 34*(1), 126-129.

[16] Van Diessen, R. & Steenbergen, J. (2002). *Long term preservation study of the DNEP Project – An overview of the results*. Amsterdam: IBM Netherlands. Retrieved January 11, 2010, from http://www-05.ibm.com/nl/dias/resource/overview.pdf

[17] World Values Survey (WVS). *Mission statement and goals of the World Values Survey Association*. Retrieved July 10, 2010, from http://www.worldvaluessurvey.org/wvs/articles/folder_published/article_base_52/files/constitution.pdf

Internet Links

[18] Afrobarometer Survey. Retrieved July 10, 2010, from http://www.afrobarometer.org/questionnaires/R4surveymanual_4feb08_FINAL.pdf

[19] European Social Survey. Retrieved July 10, 2010, from http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=23&Itemid=318

[20] International Social Survey Programme. Retrieved July 10, 2010, from http://www.issp.org/

[21] Survey of Health, Ageing and Retirement in Europe. Retrieved July 10, 2010, from http://www.share-project.org

[22] World Value Survey. Retrieved July 10, 2010, from http://www.worldvaluessurvey.com/

Further Reading

Allum, P., & Mehmut A. (2001). *Economic data dissemination what influences country performance on frequency and timeliness?* (IMF Working Paper No. 01/173). Retrieved January 18, 2010 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=880222

Doyle, P., Lane, J. I., Theeuwes, J . J. M., & Zayatz, L. V. (Eds.) (2001). *Confidentiality, disclosure, and data access: Theory and practical applications for statistical agencies*. Amsterdam: Elsevier Science.

Duncan, G. T. & Lambert, D. (1986). Disclosure-limited data dissemination. *Journal of the American Statistical Association, 81*(393), 10-18.

Inter-university Consortium for Political and Social Research (ICPSR). (2009). *The guide to social science data preparation and archiving: Best practice throughout the data life cycle* (4th ed.).Retrieved January 18, 2010 from http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

Joint UNECE/EUROSTAT Work Session on Statistical Data Confidentiality. (2009). Retrieved March 8, 2010, from http://www.unece.org/stats/documents/2009.12.confidentiality.htm

Mochmann, E. (2010). *European cooperation in social science data dissemination*. Retrieved May 02, 2010, from http://www.ifdo.org/data/data_access_conditions.html

The Dataverse Network Project. (2010). Retrieved January 18, 2010, from http://thedata.org

United Nations. (1998). *Handbook on civil registration and vital statistics systems: Policies and protocols for the release and archiving of individual records* (Studies in Methods Series F, No. 70). New York, NY: Department of Economic and Social Affairs, United Nations Statistics Division. Retrieved January 18, 2010, from http://unstats.un.org/unsd/publication/SeriesF/SeriesF_70E.pdf

United Nations Statistics Division. (2010). Retrieved January 18, 2010, from http://unstats.un.org/unsd/default.htm

Other Resources

For an example presentation of study findings, watch

Lesclingand, M. & Hertrich V. (2007). When the population is changing. A presentation of research findings in Mali. Paris, INED (CD).

The CD-ROM is available free of charge. Contact M. Lesclingand (marie.leschingand@unice.fr) or U. Herfrich (hertrich@ined.fr).