Rachel A. Orlowski, Frost Hubbard, Emily Blasczyk, and Dan Zahs, 2016

Introduction

The following guidelines detail ways in which the data collected within each country or culture in multinational, multicultural, or multiregional surveys, which we refer to as a '3MC' surveys, must be processed (i.e., coded, captur edited). Although these processing steps tend to be sequential, they may also have an iterative flow. Regarding the s 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. The flow involved in processing the survey data may also differ between paper-and-pencil (PAPI) and computer-assisted (CAPI) question In computer-assisted surveys, capturing the data, performing edit checks, and building data files should, at least part occur automatically while the data are being collected. In doing so, some effort may be eliminated. The data process effort, as well as the costs associated with that decision, should be considered when determining the mode of data co See Study Design and Organizational Structure, Instrument Technical Design, Data Collection: General Considerati Data Collection: Face-to-Face Surveys, Data Collection: Telephone Surveys, and Data Collection: Self-Administere Surveys for more details.

After processing, the data from each country can be harmonized with those from other countries (see Data Harmoni The calculation of outcome rates and statistical adjustments (i.e., missing value imputation, survey weight creation, variance estimation) can be performed, as described in these guidelines. Finally, the data should be disseminated as integrated cross-cultural dataset (see Data Dissemination). Substantive analyses can be performed on the disseminat dataset (See Statistical Analysis).

Processing and adjustment activities often are not given adequate attention. This is unfortunate, because costly error 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 . Often, only a errors are responsible for the majority of changes in the estimates . To lessen effort and minimize error, checks shou performed throughout the field period, while the respondent is still available, rather than waiting until the end of dat collection. The burden of programming and checking should not be underestimated .

These guidelines are broken down into Data Processing steps (Guidelines 1, 2, and 3) and Statistical Adjustment ste (Guidelines 4, 5, 6, and 7). The Quality and Documentation steps (Guidelines 8 and 9) are applicable to both. Please that this chapter assumes that 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 refresh needed.

↑ Ba

Guidelines

A. 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 orga for quality assessment of the survey implementation and (2) harmonized with other countries' data files in preparati statistical adjustment, dissemination, and eventually substantive research.

↑ Ba

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. Co both a summarization process and a translation process . All responses to a particular survey item need to be sum into a discrete number of categories. When the survey item is closed-ended (such as the response options in a 'Str Agree/Agree/Neither Agree nor Disagree/Disagree/Strongly Disagree' scale), the number of categories is explicit defined—five categories for a five-point scale. Any closed-ended questions ideally will have precoded response —that is, their numeric codes will have been defined prior to the start of data collection. (Following coding, furth transformation may occur to the coded data, such as reordering scales or collapsing categories with low cell coun dissemination in order to protect respondent confidentiality.) When the survey item is open-ended, the number of categories is not obvious, and should be determined via coding to the analytic purpose of that survey item. Codin translation process because the responses must be mapped to categories and non-numeric category descriptions n mapped to numeric values. It is possible to analyze non-numeric categorical data, but numeric codes are preferab because most statistical software is designed for numeric values.

Many code structures, also known as code frames, are defined during questionnaire and instrument development Instrument Technical Design); upon collecting the data, they are revisited and possibly revised. However, codes c be fully defined before data collection for some items, and, for example, some open-ended questions may be enti coded after data collection, or may have their code structures revised during data collection to account for answe do not fit into the existing code frame. Data quality, in these situations, depends partly upon the interviewer reco of the information provided by the respondent and partly upon the coder's ability to distinguish among coding ca and to assign the appropriate numeric value.

It should be noted that studies may have data sources other than questionnaires which require coding . Such sour could include visual images/recordings, audio recordings, and samples of physical materials and biomeasures (e. soil, saliva, blood).

Procedural steps

The creation of code frames for open-ended questions in some areas follows the same principles as the creation c ended questions. However, there are some important differences between the two processes. The guidelines belo divided into 1) items that apply to both open- and closed-ended questions, 2) to open-ended only, and 3) to close only. It is important to note that there are forms of questions that fall between closed and open-ended questions (e numerical open-ended questions, such as "how many times did you do X," or a question that has closed-ended re with an "Other—specify:" option).

For both closed- and open-ended questions:

1.1 Whenever possible and appropriate, take advantage of established coding schemes . This is true for both o ended and closed-ended questions, though the development of open-ended code frames is often further refin adapted for the particular research questions of the study. See Guideline 1.8 for more details.

1.2 Design each code frame to have the following attributes :

1.2.1 Unique numeric values and text labels. No number or text label should be used twice.

1.2.2 A code number for each possible response category (remember to include code numbers for item-mi data—e.g., 'Don't Know,' 'Refused,' and 'Not Applicable').

1.2.3 Mutually exclusive response categories for each variable (e.g. 'Full-time,' 'Part-time,' and 'Self-emplo not mutually exclusive).

1.2.4 The appropriate number of categories to meet the analytic purpose (see Questionnaire Design).

1.2.5 When using hierarchical code structures, have the first character represent the main coding category, subsequent characters representing subcategories . For example, the International Standard Classificati Occupations (ISCO) code is structured as 4 digits, with left-to-right as Major group, Sub-major group, group, and Unit group. The occupation 'data entry clerk' is 4132. Major group = Clerical support worke Sub-major group = General and keyboard clerks (41), Minor group = Keyboard operators (413), and U = Data entry clerk (4132) .

1.3 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. Decide how these codes will be reconciled once the data are harmonized. See also Data Harmonization.

1.4 Document codes in a data dictionary. There should be a data dictionary entry for each survey item (see Ins Technical Design for examples of a data dictionary entry). Each entry should contain the following informat

1.4.1 Variable ID, name, and label.

1.4.2 Data format.

1.4.3 Response options and associated code numbers.

1.4.4 Universe statements.

1.4.5 Interviewer and respondent instructions.

1.5 Building upon the data dictionary, develop a codebook that summarizes how the survey responses are asso with all of the data. The codebook includes metadata on the survey items, such as the question text and raw frequency of responses. This document can be used to facilitate quality control .

1.6 Test the instrument prior to data collection/data entry to catch any missing or improperly specified data. T instrument at data entry, as well as when reviewing the data produced. Sometimes a data entry application w accept a value, but the data are not stored properly. Look for:

1.6.1 Missing categories.

1.6.2 Incorrect value limits (e.g. variable on weight in pounds only accepts values 1000 or below).

1.6.3 Improperly specified data structure such as:

- Character vs. numeric field consistency.
- Field size (e.g., name field only holds 15 characters and names collected are longer than 15 charact

1.6.4 Entirely null variables, indicating instrument logic is omitting the question.

For open-ended questions:

The following example from the telephone study Survey of Consumers (SCA) 2012 will be used to illustrate con pertaining to open-ended coding :

A2. We are interested in how people are getting along financially these days. Would you say that you (and your living there) are better off or worse off financially than you were a year ago?

1. BETTER NOW
3. SAME
5. WORSE NOW
8. DK
9. NA

A2a. Why do you say so? (Are there any other reasons?)

The open-ended responses to A2a were coded into numeric categories representing reasons the respondent felt be worse off. See Appendix A for a full code frame of this example question.

1.7 There are standard code frames that are used internationally to create comparable data. These should be us survey where relevant. For example, for occupation coding there is the International Standard Classification Occupations (ISCO).

1.8 The creation of new code frames for open-ended questions is a challenging and important part of data pro it is said that "coding is analysis" . The concepts and analytic items for coding open-ended data are establish previous research, defined by the research goal, and discovered by coding the data. While code frames from studies may be used as a base, it is important to approach coding text without bias.

1.8.1 It is common to use pretest data to establish a code frame for the rest of the study. However, it is rare pretest will have enough responses to develop a fully robust frame. Often, further modifications may be necessary.

1.8.2 Some studies have a separate, later release for open-ended coded data to allow for the extra time nee processing.

1.9 Open-ended responses are converted to quantitative data by assessing the presence and/or frequency of wo phrases . These words or phrases are selected because they represent concepts that are of interest to the resea

1.9.1 For example, the open-ended response to the example A2a *"I'm better off because I got a raise this* would be coded to "10. Better pay" .

1.9.2 It is important to consider the context of the entire response, as there are ways context can affect hov a response.

- In the example A2a, the response "higher interest rates" is a code for both the 'better off' (code num and 'worse off' (code number 55) reasons (see Appendix A for the full code frame): in some contex higher interest rates would benefit a respondent, such as for investments, but in another context, hig interest rates might mean that the respondent will owe more on their loans. The entire response mus read to understand if the respondent sees higher interest rates as a benefit or a detriment.
- A respondent doesn't have an answer prepared in advance. They are thinking through their answer respond, and may discount or revise previous statements as they answer. In the above example, the respondent may have an answer to A2a such as *"Well, gas prices have gone down and that has help the cost of driving to work, but on the other hand my landlord raised the rent and my wife's hours g her job so overall we're worse off."* In this example, the respondent has discarded their 'better off' r and decided they are 'worse off.' This is less prevalent in a written open-ended response, but it can s occur there as well.

1.9.3 Multiple words or phrases may be coded under the same code. In the example, the SCA would code responses mentioning *"raise in wages or salary on present job, promotions, higher commissions, chan higher paying job (include Armed Forces induction or discharge) (Any family member who gets a raise coded 10); increased tips, bonuses"* to "10. Better Pay" .

1.9.4 At the same time, one open-ended response may have multiple codes assigned to it. For example, the response *"My wife started working when our child started kindergarten. Also, my grandmother passed May and I received some money as inheritance which helped us."* could be assigned both codes "12. M work, hence more income" and "13. Increased contributions from outside FU." If coding a response fo multiple items, the data may be structured similar to how a closed-ended "select all that apply" questio be. See Guideline 1.16.1 for more information on how data of this type are often structured.

1.9.5 Different disciplines may create different but equally valid code frames . For example, in the text *"T just no place in this country for illegal immigrants. Round them up and send those criminals back to w came from,"* a researcher interested in public policy may create the code 'immigration issues' for this re while another researcher interested in racial issues might create the code 'xenophobia.'

1.10 A good code frame starts with a good survey question. A poor survey question will result in responses tha unclear, confusing, or off-topic. When writing an open-ended question, it is important to consider:

1.10.1 Are you asking a question the respondent will understand and know the answer to?

1.10.2 Does the question need to be open-ended? If the purpose of the question is to capture specific catego interest, then an open-ended format may not be necessary.

- For example, one study may be interested in tracking major purchases, and would ask about each it separately, "1. Do you own a boat, yes or no?", "2. Do you own a second home, yes or no?", etc. A study, researching people's plans for a major purchase, may want to have it open-ended in order to items the researchers hadn't considered. In the first example, the researchers are interested in learni many people own boats and second homes whereas in the second example, the researchers are inter learning what items people want, which may be a boat or a second home.

1.10.3 See Questionnaire Design for more details on writing open-ended questions.

1.11 Ultimately, each of the coded items should themselves represent overall concepts that are of research inter example, a study (as cited in) on British Muslim girls conducted by Basit in 2003 coded interview data into major categories that clustered into 6 themes. One major theme was "identity," its subcategories being "ethn "language," and "religion." The relationship between these concepts can also be analyzed through relational .

1.12 The process of creating the code frame should be iterative. Every time a response is coded, it should be co with all those responses that have already been assigned that code. This ensures consistent coding and allow refinement of the codes. This is known as “constant comparison” .

1.12.1 This entire process should itself be repeated to refine and improve the code frame. In the second (or etc.) cycle, categories may be dropped, combined, or relabeled .

1.13 For interviewer-administered surveys, once a code frame is established, decide if the responses will be coo the field by the interviewer or by a trained coder after the case is complete.

1.13.1 These techniques can be combined: answers can be field-coded and later verified by a trained coder. can cut down on the cost of having an entirely separate and additional coding process.

1.13.2 If the coding is complex or has many categories, it is best to use a trained coder who can take the tir properly code the responses. It is important that field-coding not interrupt the 'flow' of the interview.

1.14 Consider providing users with both coded data and the raw (but de-identified) open-ended responses so th conduct their own content analysis.

For closed-ended questions:

1.15 Use consistent codes across survey items . For example:

1.15.1 A 'Strongly Agree/Agree/Neither Agree nor Disagree/Disagree/Strongly Disagree' scale would alwa the values ranging from 1 = Strongly Agree to 5 = Strongly Disagree.

1.15.2 A 'Yes/No' item would always have the values 1 = Yes and 5 = No (see Instrument Technical Desigr explanation of this coding convention).

1.15.3 Item-missing data from refusal would always have the values of 9 (or if two-digit code numbers, the of 99; etc.).

1.16 Be aware how data structure varies across survey software.

1.16.1 'Select all that apply' questions can come in variety of formats. Some software produces a variable f category and data contains a binary 'yes/no," indicating whether or not the item was selected; while oth software produces a variable for the total number of responses, with the first variable containing the va the first item mentioned, the second variable containing the value of the second item mentioned, and sc example:

Question:

Which of the following items do you own? Select all that apply.

1. Laptop

2. Cell phone

3. Tablet

Each category has a variable. Data indicates 1=Selected, 0=Not selected.

ID	CATEGORY_1 (laptop)	CATEGORY_2 (cell phone)	CATEGORY_3 (tablet)

1000	0	1	0
2000	1	1	0
3000	1	1	1

Each selection has a variable. Data indicates what survey item was selected first, second, third.

ID	SELECTION_1	SELECTION_2	SELECTION_3
1000	2=Cell phone		
2000	1=Laptop	2=Cell phone	
3000	1=Laptop	3=Tablet	2=Cell phone

1.16.2 Repeating question groups, used for asking a block of questions that repeat for distinct events/items have a variety of formats. Some software produces a wide file with repeating columns for each group, others produce a row for each event/item. For example:

Questions:

A1. Could you estimate the date of your [most/next most] recent hospitalization?

A2. What was the most immediate reason that led to your visit on [DATE]?

1. Chest pain

2. Shortness of breath/difficulty breathing

3. Physical injury (sprain, break, bleeding)

4. Other

Data structure is wide, repeating columns for each group:

id	numvisits	date_1	reason_1	date_2	reason_2	date_3	reason_3
1000	2	3/15/2015	1	12/3/2015	2		
2000	1	5/17/2015	3				
3000	3	6/21/2015	2	8/13/2015	2	11/7/2015	2

Data structure is long, repeating rows for each event/item:

id	visitnum	date	reason
1000	1	3/15/2015	1
1000	2	12/3/2015	2
2000	1	5/17/2015	3
3000	1	6/21/2015	2

3000	2	8/13/2015	2
3000	3	11/7/2015	2

1.16.3 Data may need to be transformed to meet the analytic purpose.

Lessons learned

1.1 Data are often recoded and transformed in post-processing. It is important to budget this time and expense study.

⇡ Ba

2. Decide how coding and data capture will be conducted and monitored.

Rationale

The methods used to create coded data will vary depending on several factors. One of the major factors that dete coding is the mode of data collection. All surveys require coding to classify responses. However, a paper instrum requires a separate process (data capture) to convert the physical survey into a digital data file, whereas a compu instrument may only need open-ended responses to be coded.

When using a paper-and pencil-questionnaire (PAPI), it is important to capture all data provided, even when skip are not followed correctly. Develop a protocol to handle errors when editing the data (see Guideline 3 below).

It is also important to capture information other than the survey data, such as the information from the covershee household observations, and interview details (e.g., date, time, and length of the interview), for each sample elem These data will aid in monitoring, evaluating, and potentially improving the data collection process. There are alternatives to manual keying, such as optical character recognition (ICR) (commonly known as 'scanning'), mar character recognition (MCR), voice recognition entry (VRE), and touchtone data entry (TDE).

The resources available will often dictate how data capture will be conducted. The data from all countries may b at a single location (typically the coordinating center), or it may be conducted by each country individually and c afterward .

The decisions for how coding will be monitored are also affected by these factors. Some method of monitoring i important to ensure data quality. Even computerized questionnaires require monitoring for errors.

Procedural steps

2.1 Determine how data capture will occur. This may vary across countries depending on their respective amo funding, resource availability, infrastructure constraints, and cultural feasibility. When country-specific ada are necessary, it is important to establish a data capture monitoring system that ensures comparability across countries.

2.2 Design the coding harmonization strategies needed for the data to achieve comparability across countries. more information, see Data Harmonization.

2.3 Design the data entry software to maintain the question order and measurement units of the paper survey. case of mixed-mode studies, it may also be necessary to reconcile differences between the data captured via modes. The primary goal should be to make data entry a simple and logical process, but maintaining consist between the two modes is also important.

2.3.1 For paper surveys, decide whether or not to program the software to allow the keyer to ignore errors filling out the form (e.g. when the skip pattern was not correctly followed). The decision depends on w not it is of interest to capture these errors.

2.3.2 Consider distributing a data entry shell to all study site countries that are using PAPI in a 3MC surve facilitate data harmonization.

2.4 Depending on resource availability, as well as the data being collected, consider centralized coding vs. decentralized coding. Centralized coding occurs at one location, typically the coordinating organization. Decentralized coding applies to situations where each individual country conducts its own coding prior to th being combined, as well as situations where coders from one organization work in multiple locations, such a own homes. Keep in mind that:

2.4.1 Supervisory control is easier with centralized coding. This often results in higher inter-coder reliabili Appendix B).

2.4.2 Centralized coding typically involves fewer coders, with each coder having a larger workload. The la workload can result in a higher coder design effect (see Appendix C). Training is key to reducing this e

2.4.3 Decentralized coding often occurs when administrative data such as hospital records are collected an 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 recogn that some recoding of such data may be required .

2.5 Properly train coders on the study's coding design, and periodically assess their abilities. This ensures that have equivalent coding abilities and that coding is consistent, which reduces coder design effect.

2.6 Endeavor to control manual coding by using independent, rather than dependent, verification .

2.6.1 In independent verification, two coders code all responses separately. Discrepancies are handled with computer or an adjudicator .

2.6.2 Independent verification is more costly than dependent verification, but is more reliable.

2.6.3 Independent verification reduces the likelihood of under-detection of errors.

2.6.4 Independent verification also reduces coding errors:

- The likelihood of two or three coders independently assigning the same erroneous code is small.
- However, independent verification is not foolproof, especially if the coders are not properly trained monitored.

2.6.5 In dependent verification, the first coder codes responses, and a second coder verifies them and make changes to any codes they deem erroneous, meaning the verifier has access to the initial outcome and r any detected errors.

2.6.6 A survey can use both independent and dependent verification to offset cost. Consider using indepen verification for key items that are difficult to code (such as occupation coding) and dependent verificat other items that are more straightforward, such as a 'strongly agree' to 'strongly disagree' scale.

2.6.7 Strive to verify 100% of the data entry (see and).

2.6.8 Look for the following keyer errors :

- Wrong column/field.
- Corrected/modified (misspelled) responses.
- Be especially cautious about correctly coding the first character of hierarchical code structures, bec errors at the higher levels are usually more serious.
    - For example, the code is structured as 4 digits, with left to right as Major group, Sub-major gro Minor group, and Unit group. The occupation 'data entry clerk' is 4132, whereas 5132 is the oc code for 'bartenders'.

2.7 Consider automated alternatives to key entry, including :

2.7.1 Optical character recognition (OCR) to read machine-generated characters.

2.7.2 Intelligent character recognition (ICR), commonly known as scanning, to interpret handwriting.

2.7.3 Mark character recognition (MCR) to detect markings (i.e., bubbles).

2.7.4 Voice recognition entry (VRE) to automatically transcribe oral responses.

2.7.5 Touchtone data entry (TDE) to interpret numbers pressed on a telephone keypad.

2.8 When using automated coding systems:

2.8.1 Decide between using exact matching, which results in less error but also fewer assignments, or inex matching, which has the opposite outcome.

2.8.2 Check for any responses that are left uncoded and manually code them.

2.8.3 Frequently recalibrate and configure scanning equipment to minimize the frequency with which the s misreads information (e.g., with OCR).

2.8.4 Store the code structure as a dictionary database with alternative descriptions, so a realistic response can be handled.

2.9 Evaluate the coding process.

2.9.1 For manual keying: collect and monitor paradata on coding and verification (such as error rates) at th variable, code number, and coder level.

2.9.2 For automated coding: collect paradata on the scanning operation (such as rejects and substitutes) by character and by machine.

2.9.3 Assess the reliability of coding.

- A common way to calculate reliability of a code is to compute the inter-coder reliability, or Cohen's (i.e., a statistical measure that accounts for chance). Kappa is most informative when there are a sm 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 co revising the coding frame.
- Consider revising the code if the original code is not reliable.

2.10 Flag any concerns from keyers or errors from the automated system for expert review at a later time, durin editing (see Guideline 3 below). Errors should not hinder the performance of the keyers or halt automated co

Lessons learned

2.1 Although using a comprehensive data dictionary for automated coding generally results in less manual coc expanding the dictionary does not always mean more accuracy . Additions to a data dictionary or coding ref file can lessen the automated coding software's ability to exactly match and assign code numbers to the resp resulting in more manual coding. The Canadian Census of Population and Housing in 1991 updated their ref file not only to add items, but also to remove phrases that were generating errors .

2.2 With automatic coding, consider the effort made in revising the codes in relation to the automation gained data dictionary for one of the Swedish household expenditure surveys was updated 17 times, increasing in si 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– of only 6% after 14 additional updates.

2.3 Those with prior experience coding survey data may not always be the best people to code data on a partic survey. Substantive knowledge may also be necessary when selecting coders, depending on the complexity c survey items. For example, the World Mental Health Survey employs coders who are psychologists or psych in order to diagnose verbatim responses.

2.4 Coding errors are not trivial; they can systematically alter results and damage the accuracy of estimates.

2.5 A computerized instrument does not prevent data errors. For example, if the instrument has incorrect skip has improper specification to columns, data will be lost or truncated.

2.6 Many established 3MC surveys are partly or wholly paper-and-pencil based, making data entry necessary. studies vary somewhat in the details, typically, each participating country is responsible for entering and cle own data, a supervisor or data manager checks questionnaires before data entry occurs, and some percentage questionnaires are double-entered. Whatever protocol is used, it is important to fully document the data entr process. The following are examples of data entry strategies for studies that were partially or entirely paper a pencil:

2.6.1 Round 6 of the used a paper-and-pencil instrument. Each participating country was responsible for e checking, and cleaning its own data. The project utilized a data-entry template which outlined the varia names and data types required but allowed each country to have its own questions or codes. The data w reviewed by the core partner data managers and the Afrobarometer data manager. Data cross-checks w performed on a regular basis. Either rolling data entry or batch data entry was employed at the discretio data manager. A minimum of 25% of all questionnaires was double-entered.

2.6.2 In the [Asian Barometer], another pencil-and-paper survey, quality checks are implemented at every st data cleaning involves checks for illegal and logically inconsistent values. A minimum of twenty perce data are entered twice by independent teams.

2.6.3 Round 5 of the was administered as either a pencil-and-paper or a computer-assisted survey, dependi each country's resources. National coordinators were responsible for entering and cleaning their own da documenting their cleaning procedures before submitting the data to the ESS Archive. Files were furth scrutinized for content and consistency once uploaded to the ESS Archive.

2.6.4 The is also pencil-and-paper, and each participating country is responsible for its own data editing ar cleaning. Data entry operators enter the data into a specially designed program after each of the two rou the LSMS. Each country uses computers with specially designed software to check for accuracy, consi and missing data. Further data cleaning is performed by the data manager .

2.6.5 The World Mental Health Survey can be administered as either a pencil-and-paper or a computer-ass survey, depending on each country's resources. Data from pencil-and-paper versions of the interview ar entered manually with a data entry program designed by the WMH Data Collection Coordination Cent Computer-assisted versions, by nature, are automated. Guidelines require all completed pencil-and-pap interviews to be edited for legibility, missing data, and reporting standards by specially trained editors. majority of participating countries, followups are done on questionnaires with errors. Independent doul is recommended, but keying-acceptance sampling (ranging from 10% to 20%) is allowed and used by t majority of the participating countries to evaluate keying errors. Standard coding schemes and procedu given to all participating countries. Ten percent double coding is required. Clean datasets, checked for errors such as blank or missing variables, out-of-range responses, and consistency checks, are required participating countries .

2.7 Data entry software ranges from simple spreadsheets to sophisticated applications with built-in edit check possible, a standardized set of tools should be used across countries to meet quality standards. Consider the publicly available software if cost is a concern. For instance, the U.S. Census Bureau has a data entry applic the , that is available without cost. CSPro is a software package for entering, editing, tabulating, and dissemi census or other survey data. CSPro was the recommended data entry program for the Afrobarometer Round

2.8 Sophisticated data entry software will help the staff keying the data by, for example, accounting for skip p in the questionnaire. Having this level of sophistication will likely reduce entry errors but also cost substanti more to program and to test properly.

2.9 Often, the same individual(s) creates many of the entry errors (often on the same variables). By limiting th number of individuals who perform data entry, it is easier to isolate potential problems and to offer appropri followup training.

⇡ Ba

3. Edit the data to check for errors throughout the survey lifecycle.

Rationale

Cleaning the data (i.e., correcting errors) is the primary purpose of editing, but editing can also provide informati data quality (e.g., exposing where interviewers or respondents may have difficulty performing their roles) and pc improvements to future surveys (e.g., revealing where a particular design decision may be a source of error) .

Editing can be defined as two phases: 1) identification, followed by 2) correction. Editing can occur at various pc the survey lifecycle . 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 th interviewers (or respondents, if self-administered) of inconsistent or implausible responses. This gives interviewers/respondents a chance to review, clarify, or correct their answers. Prior to data capture, survey organi can manually look for obvious errors, such as skipped questions or extraneous marks on a form. Then, during dat capture, editing software can be used to check for errors at both the variable and case level.

Procedural steps

3.1 Program computer-assisted applications to aid in the editing process during both data collection and data processing tasks. For example, in a computer-assisted personal interview (CAPI) instrument, an age value o would prompt the interviewer to confirm the value and then reenter it as perhaps 23 or 33. It may also be co 'missing' if a reasonable estimate cannot be made. See Instrument Technical Design for further discussion of instrument programming.

3.1.1 Limit programming computer-assisted data capture applications to only the most important edits so a
increase the length of the survey or to disrupt the interview/data entry .

3.1.2 Decide if the edit check is a soft check or hard check. A soft check asks for the value to be confirme
the survey progress with the original value. A hard check does not allow the survey to progress until ar
acceptable value is entered. A survey will often have both soft and hard checks. Limit the number of ha
checks to only crucial items.

3.1.3 If the interviewer/keyer chooses to retain the original value after the edit check, program the applicat
allow for a comment to be written about that decision. These comments can prevent erroneous editing.

3.2 Create editing decision rules both during and after data collection (see , , , , and). Rules can include:

3.2.1 Developing systematic protocols to resolve:

- Wild values (e.g., out-of-range responses, unspecified response categories, etc.).
- Implausible values (e.g., extremely high or low values).
- 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 th pregnancies, etc.).
- Entirely blank variables.

3.2.2 For paper-and-pencil instruments in particular, deciding how to resolve :

- Single-response variables with many response values.
- Illegible responses.
- Markings outside the response check box.
- Crossed-out, but still legible, responses.
- Added response categories (e.g., 'None,' 'Not Applicable,' 'Refused,' etc.).
- Incorrect skip patterns.

3.2.3 Comparing the current data to data from prior waves or to that from related respondents, when applic

3.2.4 Verifying the correct number of digits for numeric variables.

3.2.5 Setting a minimum number of items filled to be considered a complete interview (including item-mis
data on key variables).

3.2.6 Confirming the proper flow of skip patterns.

3.2.7 Flagging omitted or duplicated records.

3.2.8 Ensuring a unique identification number for every sample element, as well as a unique identification
for each interviewer.

3.3 Establish decision rules as to whether the potential errors should be accepted as correct, changed to anothe
or flagged for further investigation .

3.3.1 Follow up on the suspicious values only if they could seriously affect the estimates, weighing the co
logistics of recontacting the respondent .

3.4 Editing software may not be efficient in small surveys, but it is critical in large surveys .

3.5 Create a flag for indicating that a change has been made to the collected data, and keep an unedited datase addition to the corrected dataset . The latter will help decide whether the editing process adds value. If unedi are not kept, it is truly impossible to establish whether or not improvements have been made.

3.6 Assess a random sample of each interviewer's completed questionnaires by examining the captured data. F the use of skip patterns and the frequency of item-missing data to see if any interviewers need additional trai navigating the instrument or probing for complete answers.

3.7 Consider using logical imputation when appropriate:

3.7.1 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 val

3.7.2 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 drin consumed in the past week is blank, then the total number of drinks can be logically imputed by adding values from the individual beer, wine, and hard alcohol items.

3.7.3 Note that this is not a statistical technique; values are deduced through reasoning. Be aware of the da creating systematic error by using such logic.

3.8 Collect paradata on the editing process, so that it can gradually be improved and made less costly (see exa Guideline 8 and and).

Lessons learned

3.1 Overediting may delay the release of the dataset, reduce its relevance to users, and be extremely expensiv and). A lot of editing is not cost-effective. Make selective editing decisions based on the importance of the element or variable, the severity of the error, the costs of further investigation, and the effects of changes in estimates. Often, the level of detail required for any variable(s) depends strongly on the funding sources and purpose of the estimates. These considerations should be balanced with the other needs of the study. The tin money saved by implementing selective editing can be redirected to other processing steps or to other tasks survey lifecycle.

3.2 Editing must be a well-organized process; if it is not, ongoing changes to the data may actually reduce the . Identify fields involved in the most failed edits and repair them first.

⇡ Ba

B. Statistical Adjustment

Goal: To improve estimates of target population parameters based on sample survey data.

⇡ Ba

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 corre nonresponse bias. Therefore, reporting response rates and other outcome rates based on an established survey res standard is an important part of dissemination and publication (see Data Dissemination for additional discussion) Additionally, outcome rates often serve as indicators of a survey organization's general performance.

Procedural steps

4.1 Have the coordinating center provide a list of specific disposition codes and a clear description of how to classify all sample elements during the field period (using temporary disposition codes) and at the end of the period (using final disposition codes). These disposition codes will allow the standardization of outcome rat calculations across countries.

4.1.1 Generally, disposition codes identify elements as a completed interview or a non-interview. Non-inte are further subdivided depending upon whether the sample element is eligible or ineligible to participat study. For surveys where sample elements are people, ineligible non-interviews might include the respo 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 ot defined by the study.

4.1.2 Disposition codes are mutually exclusive, and while each sample element may be assigned a number different temporary disposition codes across the field period, ultimately it will be assigned *only one* fin disposition code.

4.2 Based on an established survey research standard, assign all sample elements into mutually exclusive and exhaustive categories and calculate response rates.

4.2.1 Assigning elements into predetermined final categories makes it possible to recalculate each country response rate in a standard way for comparison across countries, as appropriate.

4.2.2 The World Association for Public Opinion Research (WAPOR)/ provides one example of an establis survey research standard.

- According to WAPOR/AAPOR's "Standard Definitions of Final Dispositions of Case Codes and O Rates for Surveys," there are four main response rate components: Interviews, Non-interviews—El Non-interviews—Unknown Eligibility, and Non-interviews—Ineligible.
- WAPOR/AAPOR defines six separate response rates (RR1–RR6) :
  - Response rates ending in odd numbers (i.e., RR1, RR3, and RR5) do not consider partially-con interviews to be interviews. Response rates ending in even numbers (i.e., RR2, RR4, and RR6) 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 eligi
  - RR5 and RR6 assume that all elements of unknown eligibility are ineligible.
- Appendices D–G in Data Collection: General Considerations contain a description of disposition co templates for calculating response rates from the AAPOR.

4.3 Based on an established survey research standard, calculate other important outcome rates such as contact cooperation rate, and refusal rate.

4.3.1 There are many different industry standards available. WAPOR/AAPOR's outcome rate calculations example of one such standard . Another has been developed by Statistics Canada .

Lessons learned

4.1 Ensure that each disposition code is clearly described and reviewed during each participating country's int training. Countries may not be familiar with the specified disposition codes or the response rate terminologi another check, consider obtaining contact attempt records from each country early in the data collection peri order to ensure that all countries are correctly identifying different outcomes and understand the difference b temporary and final disposition codes. Implement all disposition codes according to the study requirements.

4.2 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 describ situation. For example, since best practice suggests allowing the sample design to differ across countries, di disposition codes regarding ineligible elements may need to be created for certain countries.

↑ Ba

5. Develop survey weights for each interviewed element on the sampling frame.

Rationale

Depending on the quality of the sampling frame, sample design, and patterns of unit nonresponse, the distributio groups of observations in a survey dataset may be quite different from the distribution in the survey population. correct for these differences, sampling statisticians create weights to reduce the sampling bias of the estimates an compensate for noncoverage and unit nonresponse. An overall survey weight for each interviewed element typic 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 dif between the weighted sample distribution and population distribution on variables that are considered to be relat outcomes. If all three adjustments are needed, the overall weight is the product of these three adjustments.

However, it is not always necessary to create all three weight adjustments when creating an overall survey weigh the adjustments only as needed; for example, if all elements had equal probabilities of selection, a base weight w be necessary. The overall survey weight would then be the product of any nonresponse adjustment and any poststratification adjustment .

Presently, the field of survey research lacks any methodology that can help develop weights for other major surv errors, such as processing and measurement error. At this time, evaluation methods are used instead of developm application of weights.

Procedural steps

5.1 If necessary, calculate the base weight for each element.

5.1.1 Each element's base weight is the inverse of the probability of the selection of the specified element all stages of selection. If necessary, calculate the nonresponse adjustment for each element.

5.2 There are many ways to calculate nonresponse adjustments. This guideline will only explain one method, uses observed response rates within selected subgroups. This method is easier to calculate than others, but a that all members within a specific subgroup have the same propensity of responding. For information on oth nonresponse adjustment methods, see , , and .

5.2.1 Compute response rates for mutually exclusive and exhaustive subgroups in the sample that are relat statistic of interest.

5.2.2 The inverse of a subgroup's response rate is the nonresponse weight for each eligible, sampled elem subgroup.

5.3 If necessary, calculate the poststratification adjustment.

5.3.1 Multiply to obtain a weight that adjusts for both unequal selection probabilities and sample nonresp each eligible element.

5.3.2 Using this weight, calculate a weighted sample distribution for certain variables related to the statisti interest where the population distribution is known (e.g., race and sex). See for a method of computing poststratification weights when the population distribution is unknown for certain subgroups (e.g., usin or iterative proportional fitting).

5.3.3 In 3MC surveys, make sure that the official statistics used by each participating country to estimate t population distribution have the same level of accuracy. If that is not the case, seek corrections or alter

5.3.4 Divide the known population count or proportion in each poststratum by the weighted sample count proportion to compute \(adj_{ps}\).

- For example: according to 2007 estimates from Statistics South Africa, women comprised 52.2% o total population residing in the Eastern Cape Province. Imagine the weighted estimate of the propoı women in the Eastern Cape from a small local survey after nonresponse adjustments was 54.8%. Tl poststratification adjustment, \(adj_{ps}\), for female respondents in the Eastern Cape would be .52 = .953.

5.3.5 Note that missing values for any variable needed for poststratification adjustments should be impute Guideline 6 for information on imputation).

5.4 Multiply the needed weight adjustments together to determine an overall weight for each element on the d

5.5 If necessary, trim the weights to reduce sampling variance.

5.5.1 Survey statisticians trim weights by limiting the range of the weights to specified upper and lower bo (e.g., using no less than the 10th percentile and no more than the 90th percentile of the original weight distribution).

5.5.2 Trimming of weights produces a reduction in sampling variance but might increase the mean square

5.6 If necessary, consider other weight components besides the base weight, nonresponse adjustment, and poststratification adjustment.

5.6.1 There may be weight components other than the three described in this guideline, including country- adjustments and weights that account for differential probability of selection for certain questionnaire s

5.7 Apply the final weight to each record when calculating the statistic of interest.

5.7.1 Weights can be scaled for different analytical purposes. One common technique is to scale the weigh they sum to the total size of the population.

5.8 Understand the advantages and disadvantages of weighting.

5.8.1 Weighting can reduce coverage bias, nonresponse bias, and sampling bias at the country or study lev depending on whether the weights were designed to reflect the population of a specific country or the e study.

5.8.2 Caveats:

- Weighting can increase sampling variance. See Appendix D for a rudimentary measure of the incre sampling variance due to weighting.
- When forming nonresponse adjustment classes, it is assumed that respondents and nonrespondents 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 count comparability across countries can be hampered . In addition, if the poststratification adjustments d dramatically impact the survey estimates, consider not using the adjustment.

Lessons learned

5.1 Ensure that all participating countries thoroughly document their sampling procedures and selection proba at every stage of selection. Countries that do not routinely employ survey weights or use complex survey de may not be accustomed to recording and maintaining this information, and without it, it can be very difficult recreate base weights once data collection is complete.

5.2 discuss the following four properties of weights which can be used as indicators of their quality and the q the sample: mean, standard deviation, minimum, and maximum. Based on the analysis of weights from 22 s survey projects, they conclude that the overall quality of weights has improved over time despite some flaws single-sample studies regarding weights not being checked, trimmed, or rescaled.

⇡ Ba

6. Consider using single or multiple imputation to compensate for item-missing data.

Item-missing data are common in social science research data. Imputation is often used to address this problem. ' of imputation is to reduce the bias in the estimate of the statistic of interest caused by item-missing data and to pr rectangular dataset without gaps 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 .

Single Imputation Methods

Rationale

Single imputation involves replacing each missing item with a single value based on the distribution of the non-n data or using auxiliary data . It is the easier of the two imputation methods. There are several common methods, ' are discussed below.

Procedural steps

6.1 Select one of the single imputation methods available. Consider the following:

6.1.2 Overall mean value hot-deck imputation.

- Replace the missing values for a variable with the mean value for that variable across the entire dat
- While this is a very simple method to use, it can distort the distribution of the variable with imputed by creating a spike in the distribution at the mean value, potentially biasing the results.

6.1.2 Overall mean value cold-deck imputation.

- Replace the missing values for a variable with the mean value for that variable from an external sou dataset.

6.1.3 Sequential hot-deck imputation.

- Sort the dataset by specific, observed variables related to the statistic of interest. For example, imag statistic of interest is the average yearly personal income in Spain. Assume that it is known from pr studies that the yearly personal income in Spain is related to years of education and age. The datase first be sorted by years of formal education, and then by respondent age.

- See if the first element on the sorted dataset has a value for the variable that is to be imputed; in the 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 sampl elements that do have data on the statistic of interest.
- If the first element does have a value, keep this reported value and move to the second element. Th 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 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 (below) because r fitting is necessary and it has fewer complexities. Thus, sequential hot-deck imputation is more eas understood by analysts and can reduce variance and nonresponse bias.

6.1.4 Regression imputation.

- Carefully create a regression model for a specific variable that predicts the value of that variable ba other observed variables in the dataset. For example, one could create a regression model that predi number of doctor visits in the past year based on demographics such as age, sex, race, education, ar 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.

6.2 For all variables for which at least one value was imputed, create imputation flag fields that indicate whicl for each record on the data file were imputed.

Multiple Imputation Methods

Rationale

The goal of multiple imputation is to account for the decreased variance imputed values have compared to obser values. Multiple imputed values and datasets are created for each missing value. Variation in the estimates across 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 . An added level of ex is needed to perform multiple imputation, which may result in a more expensive procedure than using single imp

Due to the statistical complexity of multiple imputation methods, only the most commonly used method—seque regression imputation—is briefly described below (see and for additional detail). Please refer to for information methods.

Procedural steps

6.3 Select a multiple imputation method; consider sequential regression imputation.

6.3.1 Create multiple datasets where each missing element is based on a different trial run of a regression for each imputed item.

- This is an iterative process where one item is imputed using an imputation model, and then the next 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 mo

6.3.2 Several statistical software packages are capable of multiple imputation. , a package developed at the University of Michigan and available to users for free, is an example of one such package. R programs perform multiple imputation are also available .

6.3.3 Use sequential regression imputation when records contain different numbers of missing items.

6.3.4 Although sequential regression imputation accounts for the increased uncertainty of imputed values, time-consuming for large surveys.

Lessons learned

6.1 Researchers who employ case deletion are frequently forced to collapse regions together in order to have cases to analyze. By imputing data, regional distinctions can be maintained .

6.2 Sampling statisticians advise users to avoid imputing attitudinal variables, since attitudes can easily chang time and missing data patterns can be difficult, if not impossible, to predict. Imputation models for factual v are generally easier to specify, because they are more static and outside validation can be provided.

6.3 If item nonresponse is missing at random (MAR) given the covariates used in the imputation process, imp reduces bias, sometimes significantly. In MAR, the process causing missing values can be explained either b variables in the model or by variables from auxiliary data. (See Appendix E for more information about assu for missing data).

6.4 Imputed data are synthetic data. Computed variances using single-imputed data methods will be smaller th true underlying variances that would have occurred of a same sized sample without any missing data.

6.5 Data analysts must be able to identify real values and imputed values. Therefore, the imputation must be thoroughly documented.

6.6 Imputation procedures can vary across survey topics and populations. Therefore, different procedures may be implemented and documented within different countries, etc. For an example, see .

6.7 Even with the continual improvements in statistical software, multiple imputation methods may be hard to many 3MC surveys because it takes a greater skill level, and often more time and money, than single imputa addition, each variable requires specific treatments and evaluation on how to impute the missing values.

6.8 Check that the imputation model fits the data correctly and is well specified. A poor imputation model can increase the bias of the estimate, making it worse than not using imputation.

⇡ Ba

7. When calculating the sampling variance of a complex survey design, use a statistical software package w 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 n used in practice). Similarly, statistical software packages (e.g., STATA, SAS, and SPSS) assume simple random s without replacement, unless otherwise specified by the user. However, compared to a simple random sample desi (proportionate) stratification generally decreases sampling variance, while clustering increases it (see Sample De in-depth explanations of simple random samples, stratification, and clustering). If the correct formulas or approp statistical software procedures and commands are not applied, the calculation of the precision (i.e. sampling varia the statistic(s) of interest can be inaccurate. Therefore, analysts are cautioned to ensure they are applying the corr methods to calculate sampling variance based on the sampling design. Always compare results with the default s random sample selection assumptions to check for inconsistencies that might occur due to defective estimators.

Procedural steps

7.1 In order to use Taylor Series variance estimation, which many statistical software packages use as a defau survey data file must include at a minimum a final survey weight, a stratum identifier, and a sampling unit id for each responding sample element . The chosen statistical software package must have the capacity to acco survey weights, stratification, and sampling units in the estimation process .

7.1.1 If the complex survey design used clustering, the survey data should also include cluster identifiers f responding sample element.

7.1.2 In order to estimate the sampling variance within a stratum, at least two selections must be made wit stratum. For a sampling design that selects only one primary sampling unit (PSU) per stratum, the samp variance cannot be estimated without bias. In 'one-PSU-per-stratum' designs, the PSUs are arranged aft collection into a set of sampling error computational units (SECUs) that can be grouped into pairs for t purpose of estimating approximate variances. If a participating country uses a sample design that select one PSU per stratum, the survey data must include the SECU of each element to make variance estima possible.

7.2 When a survey data file is supplied with a series of replicate weights plus the final survey weight, balance repeated replication or jackknife repeated replication could be used to estimate variances (see Appendix F).

7.3 When estimating means and variances with statistical software packages, use the appropriate procedures a commands to account for the complex survey data. For example, SAS version 9.1.3 features the SURVEYF SURVEYMEANS procedures with strata and cluster commands to account for complex survey designs.

Lessons learned

7.1 Not all countries may have access to statistical software packages or skilled personnel. Therefore, it may b necessary to arrange for reduced fees or for centralized analysis. Alternatively, consider using free open-sou software such as R.

⇡ Ba

C. 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 organ must continue to implement quality measures to help reduce or eliminate any errors that could arise during the pr and adjustment procedures discussed above. If the emphasis on quality is relaxed during these latter activities, al time and money spent on maintaining quality during the previous tasks of the survey lifecycle will be compromis

Procedural steps

8.1 Continually monitor coding activities such as the number of responses that were coded automatically or cl after data dictionary updates .

8.2 Use data entry tools to perform keying quality checks. Have human analysts check for representativeness outliers .

8.3 Monitor editing using some key process statistics . Examples are as follows (where objects can refer to fie characters, or records):

8.3.1 *Edit failure rate = # of objects with edit failures / # of objects edited* (estimate of amount of verificat

8.3.2 *Recontact rate = # of recontacts / # of objects edited* (estimate of number of recontacts).

8.3.3 *Correction rate = # of objects corrected / # of objects edited* (estimate of the effect of the editing pro

8.4 Remove any identifying information from the production data. For example, remove any names and addre attached to each responding element or unit. (For more information, see Ethical Considerations and Data Dissemination).

8.5 When possible, use paradata and other auxiliary data (e.g., census or population files) for post-survey adju and to enhance the precision of the survey estimates. For example, population files could be used to create nonresponse weighting adjustment categories. However, in 3MC surveys, be aware of very different levels c accuracy across countries for such information.

8.6 Compare the sum of the base weights of the initially sampled elements to the count \(N\) of units on the sa frame. If the sample was selected with probabilities proportional to size, then the sum of base weights is an e of \(N\). If an equal probability sample was selected within strata or overall, then the sum of base weights sh exactly equal to \(N\).

8.7 Assign a second sampling statistician to check the post-survey adjustment methodology and the statistical syntax of the survey's primary sampling statistician. This should be done whether the statistical adjustments individually by each participating country or for all countries by a statistical team selected by the coordinati center.

Lessons learned

8.1 Make certain that all identifying information is removed from the dataset before making it publicly availal some surveys, this may require detailed geographic identifiers be removed. One survey publicly released a d that included variables which made it easy to personally identify each respondent. The principles of the Hels Declaration should be upheld (see Ethical Considerations and the).

8.2 When using official statistics for poststratification adjustments, consider the reputation of the agency. It ha suggested that some countries have manipulated official statistics. Examples of potential manipulations inclu adjustment of agricultural outputs or redefining terms such as unemployment .

↑ Ba

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 pro these different users with a clear sense of how and why the data were collected, it is critical that all properties of dataset be documented.

Documentation will help secondary data users better understand post-survey statistical adjustments that can beco intricate, such as the imputation procedures and the creation of survey weights for complex survey designs. A be understanding of these adjustments will help ensure that secondary data users correctly interpret the data. In addi post-survey documentation will indicate whether the survey organization that conducted the survey met benchma agreed to in the contract by the coordinating center and the survey organization.

Procedural steps

9.1 Document the procedures and quality indicators of the data processing. Examples include:

- 9.1.1 Data capture process.
- 9.1.2 Versions of the data dictionary and codebook.
- 9.1.3 Maintaining code files used to process data.
- 9.1.4 Training protocol and manuals for data coding, entry, and editing.
- 9.1.5 Which items were coded or recoded.
- 9.1.6 Which items were edited and their original values.
- 9.1.7 How the raw data was edited.
- 9.1.8 Who coded, entered, and edited the data.
- 9.1.9 Evaluation protocol for data coding, entry, and editing.
- 9.1.10 Measure of coding reliability (e.g., Cohen's kappa). See Appendix B for more details.
- 9.1.11 Verification protocol for coding and entry.
- 9.1.12 Data entry accuracy rate.
- 9.1.13 Protocol for editing open-ended responses (e.g., removing identifying information, correcting typogr errors, standardizing language).

9.2 If values were imputed for specific variables in the study, clearly describe the imputation method that was the post-processing documentation. In addition, for each variable where at least one value was imputed, crea imputation indicator variable that identifies whether a value was imputed for the specific variable or record i dataset.

9.3 Create a unique identification number for each sampling unit. Describe how the sample identification numbers/codes were assigned to each element.

- 9.3.1 For internal use, create and document a sample identification number for each sampling unit. It is us have components of the identifier describe the case (e.g., 0600500200101: first two digits identify the c the next three digits identify the area segment, the next three digits identify the sample replicate, the ne digits identify the household, and the final two digits indicate the order of selection of the respondents the unit, where 01=main respondent selected and 02=second respondent selected).
- 9.3.2 Create a separate unique identification number for public-use data to prevent disclosing a respondent identity. This number should contain no identifying information about responding units; it is simply a v uniquely identify a case. The identifier could maintain any structure necessary for understanding the

relationships of sample. For example, the identification numbers for members of the same household c
have the same first 4 digits.

9.3.3 Sampling frame variables that could identify respondents should be included for internal use **only** (e country two digits (06), area segment three digits (005), sample replicate three digits (002), household digits (001), respondent selected two digits (01), etc.). Sampling information can be included in public- provided it cannot be used to disclose a respondent's identity.

- For example, the sample identifier could be sensitive information if the user knew that the country Japan, and the area segment was Hokkaido. Using this information, responses to rarely-occurring s items, such as those on crime victimization, could be used to search newspaper articles and discove identity of the respondent.

9.3.4 For panel studies, endeavor to maintain the same identifiers for sampling across data collection peric both the internal and public-use data files. If this cannot be achieved, create a crosswalk table that links identifier. This is crucial for data to be comparable across collection periods.

9.4 If survey weights were generated for the study, clearly explain how each individual weight adjustment wa developed and how the final adjustment weight was calculated.

9.4.1 Each explanation should include both a written description and the formula used to calculate the wei adjustment. Below are examples of the first sentence of an explanation for different weight adjustment different countries. 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, V and Brussels.
- The poststratification adjustment factor adjusted weighted survey counts to totals from Denmark's 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 additi followup efforts, and this subsample of elements was given an extra weight adjustment factor of \(W=1/.5=2.0\).

9.4.2 If additional adjustments were used to calculate a final weight, provide a clear description of how the components were created. Examples of additional weight components include country-specific adjustm adjustments that account for differential probability of selection for certain questionnaire sections.

9.4.3 Address whether there was any trimming of the weights and, if so, the process used to do so.

9.4.4 Address whether a procedure was used for scaling of the weights (e.g., population (\(N\)), populatior (N\) in 1000s), sample size (centered)).

9.4.5 If a replicated weighting method was used (i.e., jackknife repeated replication or balanced repeated replication—see Appendix F), provide the replicate weights for variance estimation.

9.4.6 Clearly describe how each of the survey weights and adjustments should be used in data analysis.

9.5 For complex survey data, identify the cluster and stratum assignment variables made available for samplir calculations. For instance:

9.5.1 The variable that identifies the stratum to which each sample element and sample unit belongs.

9.5.2 The variable that identifies the sampling cluster to which each sample element and sample unit belor

- If the sample design has multiple stages of selection, document the variables that identify each uniq 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 san variable, i.e., a field that identifies whether a unit is in the sampling error computation unit (SECU)

9.6 If the risk of disclosing respondent identities is low, consider providing the different weight components o use datasets. However, preventing disclosure of respondent identity takes priority over providing weight components.

9.7 Discuss whether the survey met the requirements (e.g., response rates, number of interviews) outlined in t contract.

9.7.1 If the requirements were not met, provide possible reasons why the survey failed to meet these requi

Lessons learned

9.1 Innovations for Poverty Action provides a good guide to data and coding management .

9.2 The application of a unique identification code is often underestimated by survey agencies using their inte reference systems. For instance, a European survey implemented a two-year special panel survey where the conducting the study did not understand the need to link the two panel waves via one variable. Hence, the ag provided a set of hard-to-interpret 'synthetic' codes that made it difficult for users to know if they were corr analyzing the data. Much time and money were spent disentangling these codes and clarifying dubious cases

9.3 Secondary users of survey data often have a hard time understanding when and if they should use weights analyses. This issue is exacerbated in many 3MC surveys, where participating countries may apply different nonresponse and poststratification adjustment strategies. Without clear documentation of how each country their survey weights and when to use each of the weights in data analysis, the chance of secondary users eith applying or incorrectly applying weights and producing estimates that do not accurately reflect the respectiv population greatly increases. Therefore, clear documentation of the statistical adjustment processes is extren important.

9.4 A good example of how to document the key elements of the statistical adjustment process can be found ii

⇡ Ba

References

⇡ Ba