Center for Effective Global Action
University of California, Berkeley

Module 3.3: Quality Assurance and Data Management

CONTENT

1. INTRODUCTION

In the previous modules of this section, we have walked you through how to design an efficient and unbiased survey instrument, how to hire a competent team, and how to plan out the field work in a way that maximizes the chance of success.

In this module we will discuss a somewhat inter-related topic: quality assurance and data management. This concept goes hand-in-hand with hiring a qualified and trustworthy team of enumerators and managers, as well as with having a structured and detailed plan of attack to collect the data that you need. Here we will teach you about the main methods of data collection (paper vs computer based), how one can use them to check for data quality concerns, and other techniques to both preemptively and retrospectively mitigate mistakes in the field.

Experience in the field of Program and Public Policy Evaluation over the last few decades has demonstrated that quality assurance is a crucial component of every evaluation project that uses survey data. Unfortunately, this has not always been the case, and quality assurance has sometimes been neglected and deemed a rather subsidiary activity within the evaluation process. As a result, many failures can be found worldwide which a timely and comprehensive implementation of quality assurance procedures could have ameliorated.

It is still common to come across evaluation studies in which, once fieldwork has finished and data has been entered, researchers discover data problems that make analysis difficult or even impossible, such as high refusal rates, excessive missing values for important variables, omitted household members, out-of-range values, and inconsistencies between variables. Some of the most important and less evident data problems are difficult to detect before the analysis stage.

As we shall see, many of these problems cannot be solved through statistical analysis alone, and may require a revisit to the field or even a new fieldwork operation, which in most cases is temporally and financially impractical. Thus, it is not rare to find cases in which data problems have inhibited important objectives of an evaluation study or even resulted in misleading conclusions.

In order to avoid the problems mentioned above and their consequences, and in light of our experience collectively gathered over the last few decades, in this session we aim to introduce the three components of an effective and efficient quality assurance process:

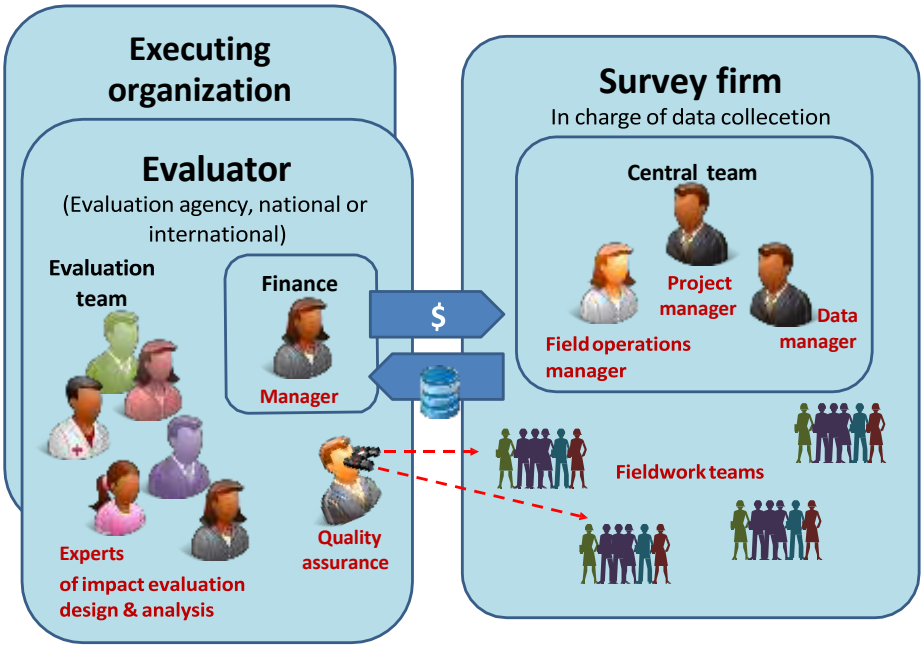
- ✓ Computer quality checks during fieldwork
- ✓ Central computer-based monitoring of fieldworkers
- ✓ Human on-site supervision

These three components of quality assurance must operate together. Implementing only one or two of them is not enough to prevent the possibility of serious data-collection errors.

2. MAIN ACTORS

At this stage, it may be useful to remember that the implementation of a survey in the context of an impact evaluation usually entails the involvement of a large number of actors:

- ✓ The organization executing the program, which is responsible for implementing the intervention being evaluate
- ✓ The evaluating organization, which may be part of the executing organization or an autonomous evaluation agency
- ✓ The firm contracted by the evaluator to conduct fieldwork, composed of:
  - The core team, formed by the project manager, the field manager and the data manager, and
  - The field teams, organized by team leader and usually including interviewers and a data entry operator. These personnel may be part of the firm's staff or may be independent contractors.



3. CENTRALIZED DATA ENTRY VS. INTEGRATION OF COMPUTER-BASED DATA QUALITY CONTROLS

As discussed in previous modules, data quality control is sometimes considered a set of tasks related to the tabulation phase of the survey, conducted towards the end of the survey project using computers at survey headquarters and generally under the control of data analysts and computer programmers. This restrictive vision of data quality control has been changing based on the experience of the past three decades, which shows that data quality control can and should play a critical role starting in the earliest stages of the survey effort. Thus, whereas many household surveys still consider data entry and data quality control as activities to be conducted in central locations once fieldwork has finished, others wisely implement the integration of computer-based data quality controls to field operations through 1) computer quality checks performed during fieldwork and 2) central computer-based monitoring of fieldworkers.

3.1 Computer Quality Checks During Fieldwork

Before the emergence of personal computers in the 1970s, centralized data entry was the only option available for transferring data from questionnaires to a computer. This method, still in use today in many surveys, considers data entry as an industrial process to be conducted in centralized workshops after the completion of fieldwork. The objective of this method is to convert the raw material (the information gathered on the paper questionnaires) into an intermediate product (an electronic file) that needs to be further refined (by means of editing programs and cleaning processes) in order to obtain the "clean" database (the final product). The priorities are speed and ensuring that the information on the files perfectly reflects the information gathered in the questionnaires. Hence, data entry operators are not expected to make take any decisions, but rather faithfully copy the data given to them.

Even though this process is almost always carried out with personal computers today, which can be programmed with multiple quality checks that can be run as data are entered, this capability is rarely used. Typically only a few quality checks are carried out during the data entry process, since data entry operators are not trained to make decisions about how to proceed when an error is detected and error detection slows down the data entry process. According to this approach, quality checks should be performed in an editing process carried out in a separate stage.

However, starting in the mid-1980s, computer-based quality checks performed as part of field operations have been identified as one of the keys to improving the quality and timeliness of household surveys. This idea was initially developed by the World Bank's Living Standards Measurement Study (LSMS) and has been applied to various other complex household surveys.

Under this strategy, data entry is performed concurrently with fieldwork through the use of computer programs that detect errors and inconsistencies while the interviewers are still on the field. In this way, errors can be corrected immediately, sometimes by means of re-visits to the household.

3.2 Problems of not Integrating Computer Quality Checks During Fieldwork

Many surveys still do not integrate data entry and quality checks into fieldwork, generally because the researchers are not aware of the advantages of integration or because of a misperception of these procedures' costs. Instead, data are entered at a later stage, normally in centralized data-entry workshops. In these cases, the central objective of data entry lays in ensuring that data files merely reflect the content of the paper questionnaires as accurately as possible. Assessing whether the content of the questionnaires is likely to reflect the reality of the surveyed households and fixing the data files when that is not the case is done in a subsequent "data cleaning" stage.

Of course, this system produces lower-quality information than that obtained through the integrated model discussed above, illuminating three shortcomings:

- ✓ The physical distance and the time period between fieldwork and data-entry ensure that it will be costly or impossible to correct errors or inconsistencies by re-visiting households. Re-visits become an exceptional procedure, frequently discussed but seldom carried out.
- ✓ A long and frustrating "data cleaning" process becomes unavoidable, which can delay the production of the database and other survey products. Moreover, the worst weakness of data cleaning is that, since it is executed in an office and without any contact with the field, it can only ensure that final data are internally consistent, remaining agnostic about whether the final data accurately reflect reality.
- ✓ Inconsistencies tend to be corrected case-by-case without any record of how decisions were made, which could lead to (and perhaps justify) users' doubts about the integrity of the data in the future.

Fortunately, post-field-work data cleaning is largely unnecessary and can be avoided if data quality checks are integrated into the fieldwork stage.

3.3 The Benefits of Integrating Computer Quality Checks into Fieldwork

In the 21st century, there is no longer any reason to use centralized data entry. Centralized data entry requires greater expense and more work. The use of computers on the field is easy and does not require expensive technology. It is a proven approach that has been successfully used for more than three decades in household surveys all over the world.

By integrating data entry into the survey fieldwork, the traditional sequence – interviews, then data entry, then data cleaning – is replaced with a continuous process with constant feedback, the outcome of which is a single final product: a database.

Integrating computer-based quality controls to fieldwork is advantageous because:

- ✓ It provides immediate feedback on field staff performance, allowing early detection of any inappropriate enumerator behavior;

- ✓ It ensures that uniform criteria are applied throughout the full data collection period and throughout the whole territory where the survey is being conducted, because the same software is being used by all of the field teams;
- ✓ It permits the correction of inconsistencies at their source by re-visiting the household if necessary, rather than having the data "cleaned" by someone sitting in an office making theoretical decisions about what should be entered. As stated above, office cleaning processes at best produce databases that are internally consistent but do not necessarily reflect the reality observed on the field;
- ✓ The integration of computer-based quality checks can generate databases that are ready for tabulation and analysis in a timely fashion, generally just a few weeks after the end of field operations. In fact, databases may sometimes be prepared simultaneous with the field work;
- ✓ The daily work to ensure data quality, backed by the integrated data entry program, can motivate all team members to reach higher standards of quality and efficiency;
- ✓ It substantially improves the survey firm's credibility with data users.

4. OPERATIVE OPTIONS FOR THE INTEGRATION OF COMPUTER QUALITY CHECKS

There are two operative options to put integrated data quality checks into place. Neither of these two alternatives is an end in itself; rather, the best operative alternative depends strictly on the specific context in which each survey is carried out.

4.1 Computer Assisted Field Edits (CAFÉ)

Computer Assisted Fieldwork Entry (CAFÉ) is a scheme that requires the field staff to be organized in teams, usually formed by a team leader, two to four interviewers, and a data entry operator. Data entry is thus, performed by each team, as we have frequently discussed in previous modules. The specific organization of field operations will depend on the technological options available for data entry.

The two most-used set-ups involve desktop or laptop or tablet computers:

- ✓ The data-entry operator of each team works with a *desktop computer* in a fixed location. Fieldwork is organized so that interviewers can visit each survey location (generally a primary sampling unit) at least twice. After the interviewers' first visit, filled-in paper questionnaires will be submitted to the team's data-entry operator, who is in a fixed location away from the visited survey location. She enters the household data gathered by the interviewers, verifies its consistency with the help of a data entry program, and will report errors to the team leader and the interviewers. The interviewers then have to pay a second visit to the households to re-elicit data from previously-problematic responses.

Each team's data entry operator works with a *laptop or tablet computer* in the field and accompanies the rest of the team on visits to survey locations. The whole team will remain in the location until all of the necessary data are entered and certified as complete and correct by the data entry program.

In both cases, the data entry and editing program have to be necessarily developed before the survey starts, and should be tested and debugged during the pilot and field tests of the survey. As we discussed in the last module, each of these procedures has specialized electricity needs that may be constraining in some locales.

CAFÉ in Iraq LSMS

In the 2007 Iraq Living Standards Measurement Study (LSMS), security issues made it impossible for data-entry operators to accompany their field teams with laptops. Instead, a data-entry office was set up in each regional capital, where data-entry operators of different teams worked in the same room.

Each operator belonged to a specific team and reported to one team leader (not to a workshop manager, as occurs in data-entry workshops). The operators entered the data from the questionnaires brought back by the interviewers of her own team and raised any matters of quality control, as they arose, with her team leader.

4.2 Computer Assisted Personal Interviewing (CAPI)

Interviewing with hand-held computers or tablets, generally referred to as Computer Assisted Personal Interviewing (CAPI or CAI), is very appealing for a number of reasons: it allows researchers to eliminate paper questionnaires; it allows for the automation of certain parts of the interviews, such as skip instructions; and it eliminates the data-entry process entirely, since data is directly entered into a database program during interviews. However, CAPI is a recently-developed technique, and there still exist numerous instrumental challenges that must to be addressed upfront.

CAPI has proven effective in linear interviews like Labor Force Surveys and the collection of prices for the consumer price index. However, when it comes to more complex surveys, there are a number of data management issues, which should be carefully addressed so that CAPI can be successfully implemented:

- ✓ The data entry program interface sometimes consists of a series of questions appearing one after the other on a screen, which is quite straightforward. However, in other cases it needs to reproduce the structure and visual format of the paper questionnaires, showing many data-entry fields at the same time. This is particularly important in modules on expenditure and consumption, where the interviewer needs to see several consumption items simultaneously. The interface must also allow for marking questions in case of doubts, and it should also make it

possible to conduct a second interview with the household without repeating the entire questionnaire. All these necessary features may result in a much more complex interface and a longer programming stage, what should be accounted for when allocating time to this stage in the work plan.

- ✓ A good CAPI application requires a highly-refined data entry program, with all the checking and consistency-validation functions integrated into fieldwork. This is of critical importance, allowing errors to be corrected during the interview itself.
- ✓ Interviewers' training is always more complex and takes longer with CAPI than it does with paper interviews, since each interviewer must be able to understand computer-generated warnings quickly and fully, and to make decisions rapidly when she is working alone.
- ✓ Finally, since CAPI interviewing needs to be supervised just like other interviewing techniques, its use should be complemented with mechanisms that allow the team leader consolidate the information from the interviewers in her team, so that she can re-visit a random sample of the households that have been visited by each interviewer and re-ask some of the questions while the team is still in the sampling unit.

Finally, the social contexts in which surveys are conducted usually do not allow for the use of high-tech expensive machinery. The use of hand-held computers or notebooks in the field may sometimes jeopardize the interviewer's safety, prohibiting their use in those cases.

In sum, both CAFÉ and CAPI share the benefits of integrating data quality checks into fieldwork, but neither is a clearly superior method. Both have strengths and weaknesses, and the decision of which to implement depends on particular contexts and situations: time available for programming and training, human resources available for the development of data entry programs, interviewer's experience, etc. There are many cases in which CAFÉ was applied in situations in which CAPI could have made work proceed easier and faster, but there are also many cases where the use of CAPI led to surveys' failure on technical grounds, where CAFÉ would have generated analyzable result.

5. TYPES OF DATA QUALITY CHECKS

Regardless of strategy, data must be subjected to five kinds of checks while being entered: range checks, checks against reference data, skip checks, consistency checks, and typographic checks. Here, we review the nature of these checks and the ways they can be implemented under the various operational set-ups.

5.1 Range Checks

Range Checks are intended to ensure that every variable in the survey contains only data within the domain of valid responses. Categorical variables should take on only pre-defined values. For example, for a yes/no question, the only acceptable codes should be "1" (yes) and "2" (no). Any other value should be

flagged as an error. Chronological variables should contain valid dates. Numeric variables should lie between prescribed minimum and maximum values.

It is generally convenient to detect and correct range errors in the initial data entry phase, because range errors often result from improper data entry rather than respondent mistakes. An error flag, such as a beep and a flashing field, may be set off when an out-of-range value is entered. If the error is merely typographical, the data entry operator can correct it immediately. However, the data entry program should allow the operator to enter an out-of-range value if it correctly reflects the response and is not due to a typographical error. In that case, a written error report should be generated so the team leader and interviewer can verify the value in a re-visit to the household.

5.2 Reference Tables

A special case of range checking occurs when the data from two or more closely related fields can be checked against external reference tables. One common situation is the following:

- ✓ Consistency of anthropometric data. In this case, the recorded values for height, weight and age are checked against the World Health Organization's standard reference tables. Any value for the standard indicators (height-for-age, weight-for-age and weight-for-height) that falls beyond three standard deviations from the norm should be flagged as a possible error so that the measurement can be repeated.

5.3 Skip Checks

These verify whether the interviewer correctly followed the instructions in the questionnaire to skip certain questions depending on the characteristics of the interviewee. For example, a simple check verifies that questions restricted to schoolchildren are not recorded for a child who responded negatively to an earlier question about school enrolment.

A more complicated check would verify that the right modules of the questionnaire have been filled in for each respondent. Depending on his or her age and gender, each member of the household is supposed to answer (or skip) specific sections of the questionnaire. For instance, children less than 5 years of age should often be measured in the anthropometric section, but the questions about occupation are not asked to them.

It is important to note that the data entry program should not follow the skips in the questionnaire automatically. In the case of CAFÉ, the operator should instead be able to enter exactly what is recorded on the questionnaire, even if there are inconsistencies, since it is not the job of the data entry programmer, or the operator, to decide how to resolve inconsistencies. For example, if the answer "No" is entered to the question "Are you enrolled in school?", the fields in which to enter data about the kind of school attended, grade in school and so on, should still be presented to the data entry operator. If

there are answers actually recorded on the questionnaire, the program should flag a potentially-incorrect skip. The team leader or the interviewer (or the centralized editing clerical staff) can determine the nature of the mistake at a later time, since it may be that the "No" was erroneously recorded in the first question. If the data entry program had automatically skipped the following fields, the error would not have been detected or remedied. The situation is subtler with CAPI, because automatic skipping is a very tempting option for the programmer, but the underlying risks are the same. This adds nuance to the development of CAPI applications.

5.4 Consistency Checks

These checks verify that responses from one question are consistent with those from another question. A simple check occurs when both values are from the same statistical unit: for example, the date of birth and age of a given individual. More complicated consistency checks involve comparing information from two or more different units of observation.

No natural limit on the number of consistency checks exists. Well-written versions of the data entry program for a complex household survey may have hundreds of them. In general, the more proper checks, the higher the quality of the final dataset. However, given that the available time to write the data entry program is always limited (usually to about two months), some expertise and good judgment are required to decide exactly which consistency checks should be prioritized.

Certain consistency checks that are applicable in almost all household surveys have proved to be particularly effective, and thus have become a de facto standard. These include:

- ✓ Demographic consistency of the household. The consistency between the ages and genders of all household members is checked with a view to kinship relationships. For example, parents should be at least (around) 15 years older than their children.
- ✓ Consistency of age and other individual characteristics. It is possible to check that the age of each person is consistent with personal characteristics such as marital status, grade of current enrollment (for children currently in school) or last grade obtained (for those who have dropped out). For example, an 8-year-old child should probably not be in a grade higher than third.
- ✓ Balance of income-expenditure and other indirect checks. When surveys measure consumption, income, and expenditure, it is always possible to model an income-expenditure balance. The various expenditure headings for the whole household are transformed into their monthly equivalent, with a similar transformation performed for income, and the two figures are compared. The tendency to under-declare income is well-known, and so it is usually accepted that the income-to-expenditure ratio is reasonable if it is within the range of 0.5 to 4.0. If the figure is lower, perhaps the interviewer did not obtain sufficient information about income; however, an error of measurement unit, quantity or price paid in one of the responses is more likely. When the figure is greater than the upper limit, this check suggests that some household expenditure has been omitted. It should not

be considered as an absolute error, but merely as an indicator that there might only be partial information for that household, which should be investigated in a second visit.

The proportion of expenditure on food is another indirect indicator of possible omissions or inaccuracies in the data. In some cases, using a range of 0.2 to 0.8 has helped to detect data errors. Of course, these ranges must be formulated carefully, since it is easy to write checks with a to-high Type II error rate (too many false negatives).

In addition, several other different consistency checks are possible. For instance, only households with one or more children in school should have positive household consumption for school books and schooling fees. Likewise, only households that have electrical service should report expenditures on electricity.

5.5 Typographical Checks

In the early years of survey data processing, checking for typographical errors was the only quality control conducted at the time of data entry. A typical typographical error consists in the transposition of digits (like entering “14” rather than “41”) in a numerical input.

Since the resolution of inconsistencies on different units of observation often requires going back to the household (or at least a thoughtful perusal of the questionnaire), a written report should be generated for the team leader and the interviewer to use in this process. Once the operator has completed the data-entry process for a household, he or she then runs the inter-record checks, which produce a listing with the detected errors and inconsistencies. The data-entry operator should immediately review this listing, since some errors might have been easily-fixable typographical errors. Next, the team leader should receive the listing along with the questionnaire, in order to correct the remaining inconsistencies in a second visit to the household.

Identification and assessment of missing values and outliers should be left to analysts. There is no universally acceptable solution to these issues and their treatment is very difficult to document adequately, but solutions are often critical to the interpretation of the analysis.

6. OTHER CONSIDERATIONS

6.1 Implications on Budget and Calendar

After data entry has been completed on the field, the central survey team has a few steps to perform. First, the data manager should gather the files with household data prepared by the various data entry operators throughout the country and verify that all households from each period are included without duplication. Though a good system of household identifiers should ensure that no households are duplicated, there is still room for human error, such as entering data from one household in two different computers or reading a diskette twice at the central office.

Any further centralized data cleaning carried out after data entry is probably redundant, because so much data quality control has been conducted by fieldworkers. Before the availability of personal computers, data entry and data editing had to be conducted separately after the fieldwork was concluded. For certain simple surveys, such data editing might converge to a relatively "clean" data set after a few iterations, sometimes taking as long as a year. In complex surveys, however, implementing quality controls and the correction of the detected inconsistencies in batch operations became extremely complicated. Editing complex surveys in this way could take several years, and in extreme cases the resulting dataset - even if internally consistent - can be extremely unreliable because of the myriad of likely-undocumented decisions that had to be made along the way. Today, personal computers can perform powerful quality checks at the time of data entry, so data entry, data editing, and field work can be integrated into a single process. Thus the need for time-consuming and inaccurate post-survey data editing has been eliminated.

6.2 Software Development Platforms

The development of data entry programs is not simple, but the difficulty lies not so much in programming new applications as in customizing available tools to the specific needs of each particular survey. There are several development platforms available, such as CS/Pro, LSD, Blaise, Survey Solutions, and Open Data Kit (ODK).

It is very important that an initial version of the data entry program is ready before the survey pilot-test so that it can be tested and debugged, in order to have a final version used during the interviewers training sessions.

6.3 Central Computer-based Monitoring of Fieldworkers

Besides the computer quality checks just described, which focus on each specific interview, it is also necessary to monitor the overall work being done by all the teams together. To achieve this, researchers should centrally process all information as it is received from the field, assessing each team's quality by comparing their data to those of other teams. The central survey team should analyze these quality reports and decide whether there is need for special supervision of any particular team whose work appears to be inadequate, or whether new and more precise instructions should be issued.

6.3.1 Defining Fieldwork Quality Indicators

Here is a basic list of indicators commonly used for monitoring fieldwork through the analysis of = differences in performance among the teams:

- ✓ Missing values for important variables (so-called item nonresponse). Depending on the type of survey, these might be dates of birth, birth weight, employment situation, quantity and value of items consumed, or monthly incomes. For instance, one would expect similar averages of items

consumed per household in geographically-similar areas. Differences in these averages usually show that some teams are expending greater effort than others to obtain the elicited information, which may bias future analysis.

- ✓ Undercounting of certain profiles. Some interviewers tend to falsify some characteristics of the interviewees in order to save work and time by reducing the number of respondents who have to complete extensive additional sections in the questionnaire. For instance, if the answer to the question "Have you worked for at least one hour in the past 7 days?" is "Yes", then the interviewer may have to complete a substantial section on economic activity. Conversely, if the answer is "No", the interviewer will be allowed skip this section. The central survey team should ensure that these additional sections are not being systematically skipped by any enumerators or teams, Sizes of households interviewed over time. A steady reduction of the average number of members recorded in each household in an indication of field team fatigue and lack of adequate supervision.

- ✓ Accuracy in coding open-ended questions, which can be checked by comparing the actual words of the interviewees with official code tables. Obviously, this can only be done in cases where the interviewees' words have actually been recorded; if not, there is nothing to be done but trust that their responses have been coded accurately.

Fieldwork monitoring reports are intended to provide early warnings of inadequate behavior in any of the field teams. In surveys with uniform procedures, such as the Demographic and Health Surveys (DHS), these tables are standardized.
Other surveys may not have standard models for analyzing the quality of fieldwork, and the central team may need to specifically design the tables required for fieldwork monitoring.

These may include:

- ✓ Quantity of transactions recorded in the diary every day, in surveys where day-to-day activities must be described in a diary. This is a way of detecting undercounting in the final days of keeping the diary.
- ✓ Changes over time in average household size. Interviewers tend to omit household members, and record smaller households, in the final months of the survey than they did at the beginning of the operation.
- ✓ Excessive deviations in children's weight and height. When a team consistently has a higher frequency than the others of suspicious weights and heights, this indicates that this team is not taking sufficient care with measurement protocols.
- ✓ Concentration of responses ended in 0 or 5. Experience from previous surveys shows that the distribution of decimals can be an indicator of fieldwork quality. In fact, if height and weight measurements were done carefully and accurately, a uniform distribution of decimals would be expected, i.e. each numeral (from 0 to 9) should represent about 10 per cent of the observations

of the last decimal place. However, some interviewers tend to round decimals to 0 or 5. Of course, this may not critically affect the overall results of the survey, but it could be interpreted as an indication of how carefully each interviewer is doing his job.

- ✓ Overall rate of subsisting errors per team. If the number of errors of a team in the final files is very high compared to the average, there should be extra supervision of this team so that it starts functioning normally.

2006 Yemen Integrated Survey

As described at the beginning of Module 9, the computer-based monitoring implemented in 2006 Yemen Integrated Survey detected a steep reduction of reported accidents and disease cases over time, two key indicators in this survey. Reported cases steadily declined from 1,645 accidents and 705 disease cases in the first month of fieldwork to just 693 and 464 (respectively) 11 months later (drops of 58 and 34 percent over 11 months). In fact, interviewers had begun to underreport diseases shortly after fieldwork launch, because their detection led to a lengthy module in the questionnaire, with the subsequent additional burden of work.

Unfortunately, this analysis was carried out in 2007, once the survey was concluded, so there was nothing to be done. Consequently, these data, which were key to the evaluation of various health programs, were unusable for impact analysis since they would have led to erroneous conclusions overestimating the impacts on health conditions.

Nevertheless, in the survey's second edition launched in January 2014, these and other indicators have been monitored from the very beginning and have been tabulated at the field team and interviewer level. This enables the detection of inappropriate enumerator behaviors when they are still incipient, allowing for more targeted and frequent on-site personal supervision and reinforcement of training.

7. HUMAN ON-SITE SUPERVISION

This last component of data quality assurance does not involve the use of computing technology. It consists in the traditional direct personal supervision that the team leader must exert on the interviewers of his team. Even though this is a long-standing and absolutely necessary task to guarantee data quality, it is not always fulfilled; it is the central team's responsibility to oversee its compliance.

7.1 Supervision by Team Leaders

As discussed in previous modules, supervisors are responsible for on-site supervision. This involves interview observation, revision of completed questionnaires, and reviewing of computer-generated errors.

7.2 Check-up Visits

Check-up visits deserve a special reference among team leader's tasks. In addition to observing actual interviews, one of the team leader's key functions consists in conducting check-up interviews with a random selection of 10-25% of households in the sample. The check-up interview needs to take no longer than 15 minutes, and should repeat parts of the original interview to check the enumerator's recorded responses. Interviewers should be aware from the start that some check-up interviews will take place, although of course they won't know which households will be re-visited.

This procedure helps maintain high standards among the interviewers and may also help in identifying unsatisfactory interviewers. This task cannot be conducted by computers, and must be completed by team leaders. Nevertheless, it is something often performed poorly (if at all) by the survey firm, and is one of the activities that should be demanded and overseen by the evaluator.

Namibia Okambilimbili Survey

In the Namibia Okambilimbili Survey, the quality assurance team hired by the evaluator selected the questions to be re-asked during the check-up visits performed by the fieldwork team leaders.

First, a 10-15% random sample was selected from among households already interviewed. Second, for each of these households a set of questions were randomly selected, assigning a higher selection probability to questions that were more likely to lead to intentional or unintentional errors from interviewers. Third, these questionnaires (of length no higher than 15 minutes) were printed and submitted to the team leaders who were on the field. In this way, every question in the questionnaire had a chance to be supervised, with a higher chance given to the trickiest questions.

7.3 The Central Team

The fieldwork plan should include structured, periodical and random supervision of fieldwork teams by the central team. Each team leader is responsible for supervising the interviewers in his or her team, but it is always possible that she will make mistakes or stray from the ethos of the mission. This is why some way of supervising the team leaders' work must also be found. From time to time, the survey field manager should visit each team and take a few check-up interview forms back to the same households for another check. These double-checks need not be frequent, but they should be random and unexpected. The central team should also base their supervision on the reports produced by the data entry program.

In addition, the field manager and the data manager should implement complementary methods for controlling fieldwork remotely. Here are some examples:

- ✓ Data recorded on the field should be sent to the central office as soon each sample point is completed, along with the questionnaires from that sample point.
- ✓ A group should be set up to judge the quality of the data received from each sample point, to detect problems of incorrect coding or unresolved inconsistencies, and to provide immediate feedback to the field teams with precise instructions on how to correct mistakes.
- ✓ Operational problems and incorrect field procedures should be continually assessed, and resolved where necessary by sending a specific supervision mission to correct problems on the field.

7.4 The Evaluator

The relation between evaluator and firm may be briefly examined from a principal-agent perspective. The evaluator (the principal) delegates fieldwork to a survey firm expecting to receive a product with certain quality standards. The firm (the agent) is generally a private firm and, as such, is motivated to maximize its profits and reduce its costs, particularly costs related to internal supervision and monitoring over its field teams. Thus, the incentives of the firm do not always align with the preferred incentives of the principal

Misaligned incentives are highly likely in this scenario, given that the two parties have different interests and there exists high information asymmetry. The firm has more information about the problems faced on the field than the evaluator, and the evaluator cannot always ensure that the firm is acting in its best interests. The most efficient way to reduce information asymmetry and prevent moral hazard is through the enactment of external supervision and monitoring schemes performed directly by the evaluator, which will likely prevent many problems during the analysis stage.

In sum, it is not recommended to leave quality assurance under only the firm's responsibility, even if it is a trustworthy and prestigious survey firm, since survey firms generally have strong incentives to disregard quality controls required by the evaluator.

Hence, the evaluator's supervision over the firm cannot exclusively consist of the examination of progress

reports or products, but needs to include explicit quality assurance activities, like visits to fieldwork teams and direct observation of interviews, central remote monitoring of data quality, follow-ups on general survey activities, periodic meetings with the firm.

All actors in the survey scenario are responsible for quality assurance. Interviewers, field team leaders, the firm's central team, and the evaluator all have specific functions to accomplish in this respect.

8. BIBLIOGRAPHY/FURTHER READING

1. Brackstone, G. (1999). Managing data quality in a statistical agency. Survey Methodology, 25, 139-149

2. Burgess, M.J. and Brierly, R. (1995). A self-directed training course for monitors of CATI operations. Operations Research and Development Division, Statistics Canada.

3. Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L. and O'Reilly, J. (eds.) (1998). Computer Assisted Survey Information Collection. Wiley, New York.

4. Dielman, L. and Couper, M.P. (1995). Data quality in a CAPI survey: keying errors. Journal of Official Statistics, 11, 141-146.

5. Dufour, J. (1996). Labour Force Survey data quality. Statistics Canada, Methodology Branch Working Paper No. HSMD-96-002E/F.

6. Dufour, J., Kaushal, R., Clark, C. and Bench, J. (1995). Converting the Labour Force Survey to computer-assisted interviewing. Statistics Canada, Methodology Branch Working Paper No. HSMD-95-009E.

7. Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. and Trewin, D. (eds.) (1997). Survey Measurement and Process Quality, Wiley, New York.

8. Mudryk, W., Bougie, B. and Xie, H. (2002). Some guidelines for data analysis in quality control. Statistics Canada technical report.

9. Mudryk, W., Burgess, M.J. and Xiao, P. (1996). Quality control of CATI operations in Statistics Canada. Proceedings of the Section on Survey Research Methods, American Statistical Association, 150-159.

10. Mudryk, W., Croal, J. and Bougie, B. (1994). Generalized Data Collection and Capture (DC2): Release 2.5.1, Sample Verification (SV). Statistics Canada technical report.

11. Muñoz, Juan. 2005. "A Guide for Data Management of Household Surveys." In Household Sample Surveys in Developing and Transition Countries, chapter 15. New York: United Nations. UN (United Nations). 2005.

12. Stecklov, Guy, and Alex Weinreb. Improving the quality of data and impact-evaluation studies in developing countries. Inter-American Development Bank, 2010.

13. Williams, K., Denyes, C., March, M. and Mudryk, W. (1996). Quality measurement in survey processing. Proceedings of Symposium 96: Nonsampling Errors, Statistics Canada, 119-128.