



ISSN 2278 – 0211 (Online)

Data Quality Issues in a Sampling Survey

Dr. Nwizugbe Obiageri Ezenwa

Senior Monitoring and Evaluation Advisor, JSI Research and Training Institute Inc.
TSHIP Project, USAID Program, Bauchi. Nigeria

Abstract:

Getting data that can be said to be simple, meaningful, accurate, reliable and timely (SMART) is the uphill task of a good population based researcher. How to get a smart, complete and unbiased data is a harder task to accomplish. The good news however, is that it is doable if the right techniques are applied. The right approach is to strictly adhere to the dos and don'ts rule governing the planning and fieldwork processes of data collection. In order to ensure that data has integrity, the monitoring, evaluation and research officer must begin the process from the data collection planning stage. In a survey research and indeed any scientific study, there are seven stages that the research process must pass through; from the detection of problem to application of solution. This is called the stages of research. They include the identification of problem, the planning, the selection of goal, the selection of methodology, collect data with trained interviewers, analyze data and act on data.

Each of the seven steps discussed so far has some data quality implications involved, if you must get your goal right. Neglecting all or some key ones, may ruin the efficacy of what would have been some good findings. Either in the field of purely physical sciences or social sciences, once the methodology presented in a study is faulty, the entire findings cannot be valid and the recommendations are null and void.

Researchers have ethical obligations to take into account when conducting study involving the human elements. These include Non-violation of personal rights, security of the participants, and confidentiality of information and getting informed consent from a respondent among others. Beyond ethical consideration, for a researcher to get unbiased responses from the respondents, there the data collectors must observe some basic socio-cultural codes, which you can describe as norms of data collection. They include proper protocol review, equal access to information by the participants, courtesy visit to significant persons, advocacy visit to deserving authority, observe local greeting ethics and dressing code, questioning well, listening well and recording well. Not listening attentively is a weakness. It is a sign of impatience. Remember the saying "a patient dog eats the fattest bone". Inpatient persons often lose out in a contest. Inpatient interviewers may even turn off people who have valuable information to provide; and one turn-off may discourage the respondent from continuing the responses with even more valuable information.

1. Introduction

Getting data that can be said to be simple, meaningful, accurate, reliable and timely (SMART) is the uphill task of a good population based researcher. How to get a smart, complete and unbiased data is even a harder task to accomplish. The good news however, is that it is doable if the right techniques are applied. The right approach is to strictly adhere to the dos and don'ts rule governing the planning and fieldwork processes of data collection. Data can be said to be of high quality if it fulfills the "smart" rule and above all, complete. If any data is certified to be of high quality, it can also be described as data with integrity. This is what every monitoring, evaluation and research officer wants.

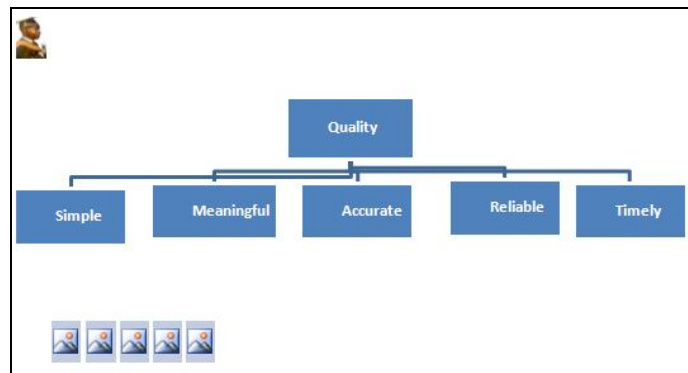


Figure 1

In order to ensure that data has integrity, the monitoring, evaluation and research officer must begin the process from the data collection planning stage. During the round table planning process, all other quality issues concerning the data collection can be discussed. The main consideration during this stage is to ensure information collection is consistent with the research ethics in Social and Behavioural sciences, and complies with freedom of information and privacy protection legislation (Guidance for Institutional Review Boards, 2001).¹In compliance with this research ethics, it is recommended that efforts be made to collect data that will shed light on issues, events or opportunities, rather than information that will violate the personal privacy of the respondent. To understand this further, it is recommended that every researcher undertakes a course on research ethics with Collaborative Institutional Training Initiative (CITI).

2. Seven Steps to Collecting Data with Integrity

In a survey research and indeed any scientific study, there are seven stages that the research process must pass through; from the detection of problem to application of solution. This is called the stages of research. Some authorities may say they are six stages, but what they often ignore is that identifying the problem itself is a critical stage. This stage should never be ignored.

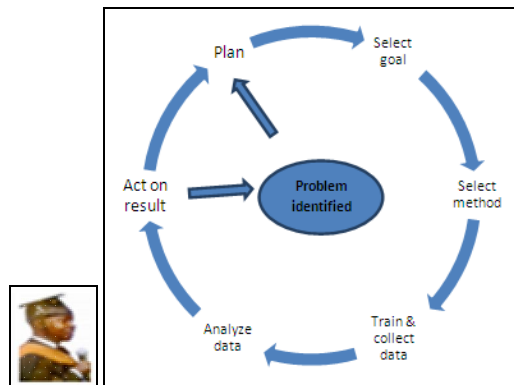


Figure 2

- **Step 1: Identify the problem**

The first step is to identify issues or the problem compelling you to look for data in order to resolve it. To do this, it may be helpful to conduct an internal and external assessment (preliminary investigation) to understand what is happening inside and outside of your organization.

- **Step 2: Plan to understand the problem:**

Get experts together to brain storm on how to go about collect the data in order to resolve the identified problem. Inputs required resources available, logistics to etc.

- **Step 3: Set your goal**

The organization defines the objectives, the purpose and goal. Focus is defined.

- **Step 4: Select a method**

An organization makes decision about survey methodology (Qualitative Data, Quantitative Data), target population, location of the survey, data selection technique, the duration of the data collection project, the tools required, among other questions.

- **Step 5: Collect data**

Visit the data collection locations and collecting data from household to household point to point, as planned.

- **Step 6: Analyze and interpret data**

Whether quantitative and/or qualitative methods of gathering data are used, the analysis can be complex, or less so, depending on the methods used and the amount of data collected. This can be done manually or electronically depending on the data analysis plan made.

- **Step 7: Act on result**

Once an organization has analyzed and interpreted the results of the data collected, it may decide to act on the data, collect more of the same type of data or modify its approach.

Quantitative and qualitative information can provide a solid basis for creating an effective action plan designed to achieve strategic organizational human resources, human rights, equity and diversity goals identified through the data collection process.

In the process of acting on your results, new problems are identified and new planning begins.

Besides new problems coming up, sometime, an organization may feel the need to expand its objectives, review its targets, or strategize in order to reach more beneficiaries. This is re-planning. They then go back to the drawing board. In some cases, an organization may decide that it needs to collect more information because there are gaps in the data collected, or areas where the data is unclear or inconclusive. This may prompt them to conduct a more detailed internal and external assessment (go back to Step 1) or try another approach. (Dillman, 1978)² This completes the data collection circle.

3. Guidelines to Quality Data Collection

Each of the seven steps discussed so far has some data quality implications involved, if you must get your goal right. These guidelines must be taken seriously from the planning stage to the data collection and data analysis stages. Neglecting all or some key ones, may ruin the efficacy of what would have been some good findings. Either in the field of purely physical sciences or social sciences, once the methodology presented in a study is faulty, the entire findings cannot be valid and the recommendations are null and void (Groves et al 2004)³. That study is only good for the trash can.

Researchers have ethical obligations to take into account when conducting interviews (one-on-one or face-to-face), case studies (individual, group, or event), focus groups (6-10 people), unobtrusive measures (artifacts, things left behind), historiography (follows a person's life history), or observations (ethnography). Researchers must use a systematic process to collect data without interfering or harming their subjects. Below is a list of ETHICAL concerns to take into consideration when performing research on human subjects (Guidance for Institutional Review Boards, 2001 *ibid*)⁴.

4. Ethical Concerns on Human Subjects

- **Non-violation of personal rights:** Privacy is one of the most important aspects with regard to ethics. Never should a researcher breach anyone's privacy. Another ethical concern is misrepresentation. Researchers are not to suggest false interest, say one thing but study another, or mask the identities of your subjects. Under no circumstance are they to ever harm or distress, (physically or psychologically) their research subjects. Throughout their research process they must leave their bias thoughts out. Whether it is subjective or objective it does not belong in the research, they must only include the facts. And finally, there is the danger of invisibility. Researchers should never ever put their subjects in a compromising position where there is a potential for danger.
- **Security of the participants:** Personal data should be reasonably guarded against risks such as loss, unauthorized access, modification, or disclosure. Environments that pose threat to life, risk of injury and psychological insecurity must be avoided. Environmental scanning is part of the preliminary planning for the research during a data collection activity.
- **Confidentiality:** Closely following ethical concern for the protection of personal right is the concern for confidentiality. A researcher may be privileged to have the knowledge of respondent secret information. This must not under any condition be revealed to the public. The respondent must be assured of protection from unauthorized disclosures of information, except among those who are expected to know (the research team).
- **Informed consent:** The respondent must not be coerced into providing information which otherwise he/she is unwilling to divulge. The respondent should expressly be told that he/she has the right to refuse a response. At all time, ask for the consent of the respondent either in writing or verbally. For the under age, their caregiver is expected to sign consent note on their behalf.

5. Other Considerations towards Achieving Quality Data

- **Protocol review:** The study or data collection protocol must be properly written, reviewed several times by a team of experts (separately and collectively), before adopting it. The protocol must not take any detail for granted. The objectives must be very clear, the methodology well explained, the location clearly delineated the target population not ambiguous, the sample size very well defined the technique for reaching the respondent well spelt out and finally the tools tested and pretested.
- **Equal access to information by the participants:** Every participant in the data collection project must be informed of any correction made to the methodology or tool. New information from the project director must be relayed to all and in all locations. Failure to do this would introduce bias into the data collection process. Supervisors must interact with the project coordinator on a daily basis for briefing on the day's events, and take an update on any new development.

- **Courtesy visit to significant person(s):** Courtesy visit is not an advocacy visit. Courtesy visit means paying homage to a deserving authority or a significant person in order to let him know you are within his area of influence. Asking for favour is not part of courtesy visit. Courtesy visit may lead to sensitization; in which case you use the opportunity to inform the host why you are in his community. It is necessary you make a courtesy visit to the key gate keeper in your location before you commence your field activities. Do not wait until you run into muddy waters before you remember that there are community gate-keepers. The time spent during courtesy visit is actually time invested into assuring oneself, or the team against the hazards of the unknown in the community.
- **Advocacy visit to deserving authority.** It is necessary to differentiate between sensitization visit and advocacy. The essence of sensitization visit is to give information, to improve awareness and to prepare people for a given situation. Advocacy begs for recognition, for assistance or support in order to move the process forward. Advocacy is not complete without an outcome or commitment (statement of intent to help or to recognize) from the host.
- **Observe local greeting ethics:** Learning basic language and mannerisms of the host communities may be very helpful during the data collection period. Greet local hosts (respondents and their household members) in their local language. Make kind gestures to them like playing with little kids found in the home, showing positive concerns for bad conditions and appreciating your environment. This creates the feeling of “we are one; we can work together, we can break barriers”
- **Dress like them:** Your dressing speaks a lot about you. If you dress like a stranger, the community will see you as such; on the other hand if you dress like them, you will be seen as one them. Dress like them, look like them, welcome them, then they will work with you.
- **Verbatim questioning:** Ask your questions exactly the way they were framed in the questionnaire. Do not attempt to interpret the questions to the respondent. Do not attempt to breakdown the questions to suit the understanding of the respondent. Any of these ways can introduce response bias. Remember, these questions have been tested and found suitable for the target respondents. If there is a need to interpret the questions into any particular local language or dialect, this ought to have been done before the field interview. Researchers must therefore first conduct a preliminary survey of the target population in order to determine their peculiar needs in the proposed study. If the population is predominantly illiterate in English language, it will be beneficial for the questionnaire to have two or multiple versions viz; English language and the dominant language versions (Tourangeau, R., and T. Yan. 2007)⁵.

5.1. Interviewers Listen

According to Nichols and Stevens (1957)⁶ listening may seem like an unimportant activity, but my experience convinces me that the reverse is true. In a survey, one of the participants said: “Frankly, I had never thought of listening as an important subject by it. But now that I am aware of it, I think that perhaps 80 percent of my work depends on my listening to someone, or on someone else listening to me.”

Not listening attentively is a weakness. It is a sign of impatience. Remember the saying “a patient dog eats the fattest bone”. Inpatient persons often lose out in a contest. Inpatient interviewers may even turn off people who have valuable information to provide; and one turn-off may discourage the respondent from continuing the responses with even more valuable information. The shift from telling to listening can be startling to respondents; and I guarantee you, respondents will receive it well.

5.2. Recording Properly

Quoting Richard Heckert (The McKinsey Quarterly, 1997 number 4)⁷; anyone seeking to be a leader should always tell the truth, if form another reason than it is simpler. He puts it this way: “If you always tell the truth, you won’t have to remember what you said.” I have observed that a high proportion of interviewers in Nigeria don’t listen very well. The consequence is that they report wrongly. If you are not patient to listen to your respondent, you may not be able to understand his/her conclusion. The conclusion is more important than the earlier statements. He may have made mistakes initially and therefore attempts to correct himself/herself subsequently, and in his/her conclusion. If you do not listen to the end, you would report him/her wrongly. What you record or report is a function of what you heard. Interviewers please record or capture correctly what the respondents’ eventually said. Do not be in a hurry to report him/her; otherwise you quote him/her out of context.

Revisit any response made by your respondent, which is not clear to you. Understand your respondent before you record a response. You can probe by asking; do you mean..? Edit your job on the spot before you move to the next respondent. Review all completed questionnaires or captured responses with your supervisor at the end of each working day, before you transmit your data to the collation centre.

5.3. Do Not Refuse Mild Offers

Do not refuse an offer of food (taste a little if you can) and say thank you; I have just eaten before coming to you. Accept an offer of fruits if given, pick a few and put inside your bag or pocket and ensure you wash them before you eat. Do not take away anything you will not eat; otherwise you deprive a needy person in the community. Do not drink bottle water or eat your food in the presence of the respondent. Do so inside your car or in your resting place.

5.4. Have Complete Coverage

- **Have a starting point:** Have a commonly agreed starting point with your supervisor before you start your coverage. This is necessary for him to be able to assess the accuracy of your sampling. A supervisor who knows your starting point would find it easier to trace you as a data collector. It is necessary that once a while, the supervisor does on the spot checking of the forms you are filling. Locating the household where you are will help him give this support, and knowing your starting point will be very helpful.
- **Leave no street uncovered:** Beginning from your starting point, move systematically round your location until you cover all streets and back to your starting point. Moving haphazardly is contra-indicated in a systematic sampling procedure. This will introduce prejudice into the sampling process.
- **Replacing an eligible respondent:** In every survey design, it is anticipated that some eligible selected elements may not be ready for interview (decline), or available for interview (absent) at the time they are expected. These two groups of respondents are always planned for in a properly designed survey. Not to anticipate them and plan for them leaves the interviewer with the option of using the rule of the thumb. This introduces respondent's selection bias.
For an eligible and selected respondent who is absent, the rule is that the interviewer drops a revisit notice, goes on with other interviews and returns back to the absent respondent to conduct the interview. This can be repeated up to three times while the interviewer is still in that community. Thereafter, if the respondent is still not available, you can declare him/her absent and therefore use an agree method to replace him/her. This is called a respondent replacement.

5.5. Errors in Data Collection

Faulty data may be due to sampling process or non-sampling error. In most surveys, there are four primary sources of non-sampling error. Here we will discuss non-sampling errors including coverage error, non-response, measurement errors, and errors that may arise during data capture and processing, and how to reduce them and achieve a clean data. (Groves, R. M. 1989)⁸

- Non-response Error.
- Coverage error.
- Measurement Error.
- Processing Error.

5.5.1. Non-Response Error

If non-response is due to the error of omission on the part of the interviewer, it becomes a non-response error. Two major types of non-response errors exist: unit nonresponsive (referring to lack of completion of any part of the survey) and item non-response (submission or participation in survey but failing to complete one or more questions of the survey).⁹⁻¹¹

Another form of non-response error is opting-out or decline interview. Failure to obtain complete data from all selected individuals due to deliberate opt-out is described as non-response due to declined interview. In survey sampling, many of the individuals identified as part of the sample may be unwilling to participate, not have the time to participate (opportunity cost)¹² or survey administrators may not have been able to contact them. In this case, there is a risk of differences between respondents and non-respondents, leading to biased estimates of population parameters. This is often addressed by improving survey design, offering incentives, and conducting follow-up studies, which make a repeated attempt to contact the unresponsive individual. The effects can also be mitigated by weighting the data when population benchmarks are available, or by imputing contacting more respondents than the minimum sample size desired.¹³

5.5.2. Coverage Error

The failure to give some units in the target population any chance of selection into the sample, or giving units more than one chance of selection. This error arises from an incomplete understanding of the population structure of a study location. Where the study element is diverse and heterogeneous, the population needs to be stratified and diverse segments represented by the sampling technique. Failure to use the right sampling technique definitely excludes some elements in the population because they were not given equal chance of representation. This introduces an error of coverage. Error of coverage becomes less significant where the study population is homogenous. This means that any sample (n) selected from the population becomes a representation of the larger population (N). See (Groves, Couper, Fowler et al 2004)¹⁴

5.5.3. Measurement Error

The inaccuracy in responses recorded on survey instruments; arising from incomplete coding, incomplete recording, omission of questionnaire and data transcription error.

5.5.4. Processing Error

Errors introduced after the data are collected, including; Data capture errors, Errors arising during coding and classification of data, and Errors arising during editing and item imputation of data.

6. References

1. Guidance for Institutional Review Boards, (2001).Clinical Investigators and Sponsors." U.S. Food and Drug Administration. U.S. Dept. of Health and Human Services. 29 Jan. 2008.
2. Dillman, D. (1978). Mail and Telephone Surveys: The Total Design Method, New York: John Wiley and Sons.
3. Groves, R. M., M. P. Couper, F. J. Fowler, J. M. Lepkowski, E. Singer, and R. Tourangeau. (2004). Survey Methodology, Hoboken, NJ: John Wiley and Sons.
4. Guidance for Institutional Review Boards, (2001 *ibid*).
5. Tourangeau, R., and T. Yan. (2007) "Sensitive questions in surveys," Psychological Bulletin, 133(5): 859–883.
6. Ralph G. Nichols and Leonard A. Stevens, "Listening to people," Harvard Business Review, September–October 1957, p. 85.
7. Richard Heckert, in THE MCKINSEY QUARTERLY (1997 NUMBER 4)7-8.
8. Groves, R. M. (1989). Survey Errors and Survey Costs, New York: John Wiley and Sons.
9. Berinsky, A. J. (2008). Survey non-response. In W. Donsbach & M. W. Traugott (Eds.), The SAGE handbook of public opinion research (pp. 309-321). Thousand Oaks, CA: Sage Publications.
10. Jump up, Dillman D. A., Eltinge J. L., Groves R. M., & Little R. J. A. (2002). Survey nonresponse in design, data collection, and analysis. In R. M. Groves, D. A.
11. American Community Survey: Report 7: Comparing Quality Measures: The American Community Survey's Three-Year Averages and Census 2000's Long Form Sample Estimates," Washington, DC. U.S.
12. Census Bureau (2002). "Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: May 2002, Report 2: Demonstrating Survey Quality," Washington DC
13. Jump up, Porter, Whitcomb, Weitzer (2004). Multiple surveys of students and survey fatigue. In S. R. Porter (Ed.), Overcoming survey research problems: Vol. 121. New directions for institutional research (pp. 63-74). San Francisco, CA: Jossey Bass.
14. Groves, R. M., M. P. Couper, F. J. Fowler, J. M. Lepkowski, E. Singer, and R. Tourangeau (2004 *ibid*). Survey Methodology, Hoboken, NJ: John Wiley and Sons