

D2.2 Conducting Randomised Sample Surveys

Author: *Heidi B. Johnston*

WHY CONDUCT A SAMPLE SURVEY?

To obtain information reflecting characteristics of a large project area population, rather than only those who participate in workshops or group meetings, one can conduct a sample survey. In a sample survey information is collected from a portion of a population to provide information meant to represent the entire population. A representative sample survey is less expensive and less time consuming to conduct than a census of the entire population, and if thoughtfully designed and carefully conducted, can yield results that accurately and reliably reflect population characteristics relevant to improving project implementation and impact.

Conducting a sample survey is a multi-step process including survey planning and design, fieldwork, data processing and analysis, and preparing and presenting findings to multiple stakeholder groups.

SURVEY PLANNING AND DESIGN

Survey planning includes developing a detailed project outline with a timetable and budget (see Table 1). The population to be studied needs to be precisely defined, key research questions need to be identified and agreed upon, the research and analysis strategies need to be decided and official permissions must be obtained.

Table 1: Example of a timeline for conducting a randomized sample survey, from Nichols (1991), p. 22.

Weeks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Planning and design														
					Fieldwork										
								Data processing and analysis							
												Presenting findings			

Questionnaires are designed with the aim of systematic collection of sets of standardized information. Applying the Questionnaire should generally take an hour or less of respondents' time; including time to introduce the survey purpose and obtain consent from the respondent to participate in the research (see example consent form, Annex 1).

Survey questions can be "closed ended" having a limited range of answers, and for which the full range of answers is included in the questionnaire form. They can also be "open ended" and depend on the respondent to provide an answer in his or her own words. Open ended questions are usually used to document attitudes and experiences for which there are many possible answers. Closed ended questions are easier to record answers to, code and analyze and should be used when possible (see example questionnaire, Annex 2). The questionnaire should be translated into the local language, using the terminology of the local community and avoiding technical terms. Back-translation of the questionnaire into the original language is useful to ensure accuracy of meaning. Pretesting of the interview forms and interview procedures is essential to confirm that the meanings of the questions are clear and not misunderstood, and that the interview procedures are thorough and make efficient use of the respondents' time. After pretesting, the forms and procedures must be revised according to the pretest results.

Determining the right sample is a multi-step process. First the target population must be defined and listed. For example, for a household survey a list could come from a map of the study area showing individual dwellings. The degree to which the sample is representative of the total population depends on the size of the sample and the method by which the sample is chosen.

A simple way to determine the necessary sample size is to first consider the information the survey is meant to yield. Consider the population subgroups, for example wealth or income groups, and key indicators such as toilet use or experience of illness. Using information already known about the community being studied or other similar communities, develop an example data table showing how the sample is expected to divide into the key subgroups (see Table 2). The total sample size should be large enough so that the sample size in each subgroup is **at least 50**, allowing for the identification of variation within subgroups. The sample should include at least 10% more than the predicted sample size to account for members of the sample who refuse to participate or are not at home. In the example shown in Table 2, the sample would be 450 plus 45, yielding a total sample of 495.

Table 2: Example table to demonstrate simple sample size calculation: toilet use by level of wealth.

Toilet use	N	Public	Private	Shared	Other
Wealth level					
I (poorest)	75	70%	0%	10%	20%
II	150	40%	10%	50%	0%
III	100	20%	60%	20%	0%
IV	75	10%	80%	10%	0%
V (wealthiest)	50	5%	95%	0%	0%
Total	450				

- A) Once the sample size is developed, the sample can be generated. One means of generating a **simple random sample** is to use a complete list of all individuals or units (such as households) in the entire population of interest (i.e. the sampling frame) and assign a number to each unit/household. Write these numbers on individual slips of paper, put these into a container and mix them well. By pulling out the number of slips estimated as necessary for the sample, the simple random sample is generated. A plan should be made to interview the units/households identified in the selection process.
- B) **Systematic random samples** are easier to select than simple random samples and are more likely to represent all subgroups of the sampling frame. Again, a complete list of all individuals or units in the sampling frame is necessary. To generate a systematic random sample from this list, select individuals or units at regular intervals from the list. For example, to attain a systematic random sample of 500 households from a sampling frame list of 2500 households, first randomly select a household from the sample frame list. From this point, choose **every 5th** household on the list, going to the beginning of the list when the end is reached. Again, after the sample is selected, a plan should be developed to interview the individuals or units identified in the selection process.
- C) **A random walk** is not an entirely random method of generating a sample as it relies on the thoroughness and judgment of a field supervisor. The main goal is to have a sample spread over the entire study area. One common random walking scheme in an area with roads or paths is to select a central junction of the area. From this point a field supervisor should send pairs of interviewers in different directions of the junction. Individuals in the pairs work on opposite sides of the road/path. For a sample of 500 within a sampling frame of 2500, interviewers should interview at every 5th household. If no one answers at a selected household the interviewer should note the household and return for an interview at a later time. On subsequent days the field team should go to smaller

junctions sufficiently far from the previous days' sites to avoid repeat contact with the same households. An area map will help the field supervisor track the structure of the sample (see sample map below).

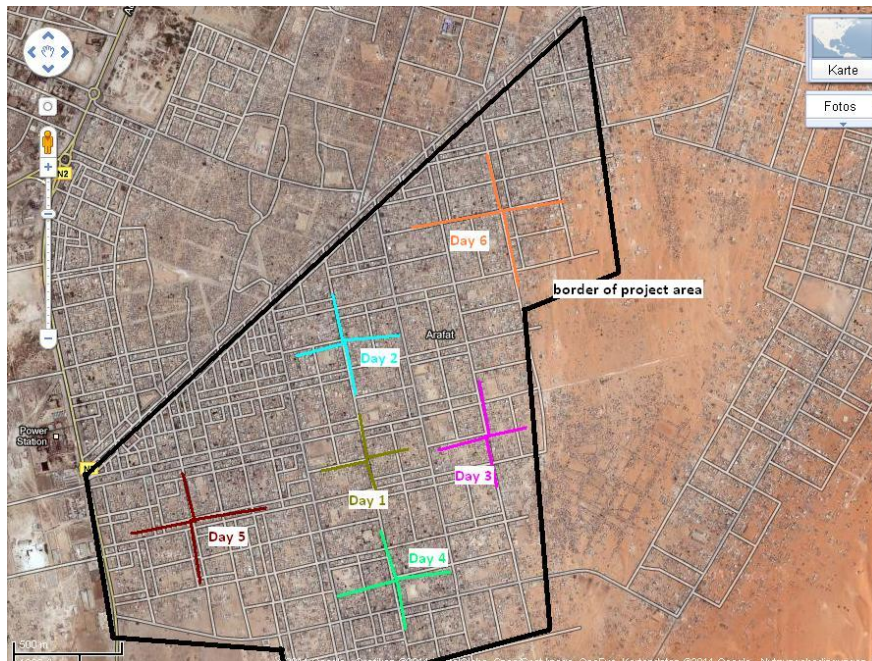


Figure 1: Random walk schedule for randomized sample survey (example from Nouakchott, Mauritania)

FIELDWORK

Selection and training of interviewers is key to the quality of survey results. The age, sex, and social status of interviewers can affect interview dynamics and should be taken into consideration during fieldworker recruitment, as should interviewing skills and experience. An interviewer training program should take a minimum of 2-3 days and cover topics such as the background and purpose of the survey, introducing the survey to respondents, obtaining permission from respondents to participate in the survey, and use of the interview form. Interviewers should repeatedly practice interview procedures and use of the questionnaire prior to the survey implementation, and prioritize making the respondent feel comfortable when answering questions. Careful supervision and management are essential to keep fieldwork running smoothly and to ensure the collection of quality data.

DATA PROCESSING AND ANALYSIS

When data collection is being completed **data processing** can begin. This begins with checking completed forms to note and correct errors. All responses to the survey

must have numerical codes to enable data entry. For closed ended responses numbers beside each response variable represent the numerical code. For open ended questions the range of responses need to be manually grouped into meaningful categories, with numerical codes assigned. Coded data can be entered into a database using data entry software.

Data collected in surveys can be analysed in different ways. Tables can be used to summarize findings; two-way tables (cross tabulations) show variables of interest according to population subgroups. Presenting data in percentages allows for comparison of subgroups. Presenting survey results graphically – for example in bar charts, line graphs, histograms and scatter diagrams – can ease the interpretation of data. It is important to analyse the distribution of variables of interest in the population, and this can be done using percentiles and standard deviations. The advice of a statistician or local university researcher in survey planning stages and prior to data analysis can be invaluable.

PRESENTING FINDINGS

Effective communication of research findings to the right audience is essential for research to have an impact. Consider the full range of audiences that need to be reached, such as project officers, community members, government officials, technicians and aid agency staff. A basic starting point is a good report describing why and how the survey was conducted, presenting findings with tables and diagrams, and listing conclusions and recommendations. In addition results may be presented in community and professional meetings, and perhaps using other communication strategies such as drama, newspapers and video.

RECOMMENDED RESOURCES

Nichols, P. (1991). Social survey methods: a field guide for development workers, Oxfam.

This text is an excellent, highly readable introduction to the multiple steps of conducting social science surveys with a focus on low resource settings. Available at: http://books.google.ch/books?id=MLqyxiTDiNoC&printsec=frontcover&dq=Nichols+social+survey+methods&hl=de&ei=K7yBTZqmBliWswav0vS2Aw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CDQQ6AEwAA#v=twopage&q&f=false.

Epi Info™ is highly recommended free software available from the United States Centers for Disease Control. Different modules of Epi Info™ facilitate the development of questionnaires, data entry forms, data entry, analysis and data reporting. Available at: <http://wwwn.cdc.gov/epiinfo/>

Annex 1: Example consent form for household sanitation survey

[Name of Institution]

Written Consent Form

Title of the Research Project:

Principal Investigator:

Purpose of the Research:

Why Selected: We are interested in learning about sanitation in your community. You are one of [#] people in this community who are being asked to answer a few questions to help us learn more.

What is expected from the respondent?

We are asking you to help us complete a short questionnaire. Using the questionnaire I will ask you to answer questions about your household. The survey will take approximately 45 minutes to complete.

Risk and benefits: There is no more than minimal risk to participating in this study. Although you will not be paid for your participation, the responses you give will provide us with valuable information to understand sanitation issues in your community. This information will be used to make recommendations to improve sanitation.

Privacy, anonymity and confidentiality: Please be assured that the information you provide will be treated in a confidential manner. Only investigators of this study and members of the Ethical Review Committee of [Institution] will have access to the information collected. Your name will not be associated with the information you provide us. Instead, this questionnaire has been assigned a unique identifying number. You will not be personally identified in any reporting of data. All information collected will be combined and reported as a group.

Future use of information: This information may be used in the future for additional research on sanitation. We assure you that privacy, anonymity, and confidentiality will be maintained.

Right not to participate and withdraw: Your participation is completely voluntary. You have the right to stop the survey at any time.

Person/s to contact: If you have any questions about your rights as a participant in this study, you may contact [principal investigator] at [institution]. His/Her contact number is [contact number]. If you feel that you have been treated unfairly or have been hurt by joining the study you may call [principal investigator] or [name] at [institution] His/Her contact number is [contact number].

Principle of compensation: You will not be paid for your participation in this study.

Do you have any questions? Yes No

Do you agree to participate in the study? Yes No

We are now inviting you to participate in our survey and provide consent. If you agree, please indicate that by providing your signature in the space indicated below.

Thank you for your cooperation.

Signature or left thumb impression of respondent Date

Annex 2: Example survey on household sanitation (47 questions)

Baseline household Questionnaire

Date: Location: Name of interviewer: Person interviewed: (sex/hh head)

Time needed: ___ minutes Interview completed? (yes/no)

Note to the interviewer: The respondent should preferably be the household head. If it is not possible to interview the household head, please interview the spouse. If the spouse is not there, interview an adult household member above 18 years of age.

1. Is it possible to interview this household?

- Yes
- No, the respondent refused to participate in the interview or was not at home. If not at home, note time to return to conduct the interview:
Time 1: _____ Time 2: _____ Time 3: _____

A. Household Composition

I will first ask a few questions about the people who belong to this household.

2. How many people currently live in this household (including children)?

3. How many adults (persons of age 18 and older) live in this household?

4. How many children living in this household are between 0 and 5 years old?

5. How many children living in this household are between 6 and 10 years old?

6. Of the children (aged 0 to 10) living in this household, how many currently go to school?

I will now ask you questions about the adults of this household (persons aged 18 and above). Please, let me know:

7. Is this household head male or female? 1 = Male
2 = Female

8. What year was he/she born? _____

9. Is he/she married? 1 = Yes
2 = No

10. What is his/her ethnicity? _____

11. Is the household head one of the primary decision makers on matters related to household finances/investments? 1 = Yes
2 = No

12. Can the household head read or write? 1 = Yes
2 = No

13. What is his/her highest level of completed education? 1 = No formal education
2 = Primary school
3 = Secondary school
4 = Post-secondary school

14. What is his/her main occupation? _____

15. What is the total average monthly household income? (local currency) _____

Now I will ask you a few questions about your house itself.

23. What are the walls of the house mainly built of?

- 1 = Earth/mud
 - 2 = Wood
 - 3 = Bricks/cement
 - 4 = Other, specify:
-

24. Does the house have electricity?

- 1 = Yes
- 2 = No

25. Does the house have its own toilet facility?

- 1 = Yes
- 2 = No

26. Is the house connected to piped water?

- 1 = Yes
- 2 = No

27. What is your household's main source of drinking water?

- 1 = Public piped
 - 2 = Shared tap
 - 3 = Private piped
 - 4 = Protected well
 - 5 = Unprotected well,
 - 6 = Water vendor
 - 7 = Other, specify:
-

C. Health Conditions

I will now ask you questions about your health and the health of this household's children under the age of 5. If you cannot answer the questions about the children, we would like to ask another adult who knows about the childrens' health.

28. Have you had diarrhea in the last 2 weeks? 1 = Yes
2 = No

29. For how many days in total have you experienced diarrhea during the last 4 weeks? _____

30. Have any children under the age of 5 living in this household had diarrhea in the last 4 weeks? 1 = Yes
2 = No

31. What do you think are the two diseases that people in your community most often experience?

01 Eye infection	09 Tuberculosis	17 Headaches
02 Yellow Fever	10 Diarrhea	18 Backache
03 Stomach Pain	11 Vomiting	19 Accident
04 Malaria	12 Dysentery	20 Chicken Pox
05 Schistosomiasis	13 Urinary Infection	21 Measles
06 Intestinal worms	14 Skin infection	22 Other – specify:
07 Typhoid	15 Hepatitis A	23 Other – specify:
08 Cholera	16 Respiratory disease	99 I do not know

37. How far is the public toilet from this house? _____

38. If you use this public toilet, why do you use it? If you do not use this public toilet, why do you not use it?

39. If you use a public toilet, how much do you pay for using the public toilet?

40. If you share a facility with other households, who paid for the building of the toilet facility you share?

41. How many people do you share the toilet facility with? _____

42. Is the toilet your household most often uses generally clean?

- 1 = Very clean
- 2 = Somewhat clean
- 3 = Somewhat dirty
- 4 = Very dirty

43. Is the toilet your household most often uses usually emptied shortly after it fills? 1= Always
2 = Most of the time
3 = Rarely
4 = Never
5 = Other, specify:

44. Who pays for the emptying of the toilet? 1 = All users
2 = Only a few households
3 = NGO or project
4 = Other, specify:

45. How comfortable or uncomfortable do you and other members of your household feel because your household has no private toilet?
1 = Very uncomfortable
2 = A little uncomfortable
3 = Somewhat comfortable
4 = Very comfortable

46. What could motivate your household to construct a private toilet facility to be used by your household only?

47. Please tell me, in your opinion, what are the main issues associated with people not using a toilet facility?

Thank you for participating in this interview, we appreciate your time and responses!