Baseline Survey

A survey is a useful tool for assessing program needs and evaluating program achievements and progress. A survey implementation is intended collect additional data from a population. They aim to gather information that is not routinely collected by the existing information systems.

The **baseline study** is the analysis and description of a situation prior to the programme against which can be assessed or comparisons made. The baseline study provides a benchmark for our programme objectives focusing mainly on water, sanitation and recommended hygiene behaviour coverage. The baseline provides the basis for monitoring and evaluation, with a follow-up study at the later point (typically at mid-term of completion of the operation) to facilitate measurement of the outcomes and impact of the programme.

Objective of this type of study in the WASH sector would be measuring coverage. Coverage is the percentage of people in any given area who know of and/or practice recommended health behaviour or who receive a particular water or sanitation service. Knowing the coverage of various health knowledge and practices helps us plan by allowing us to choose priorities. We can decide to focus our efforts on improving those health knowledge and practices that have low coverage. Over time, repeated measures of coverage show us if our efforts are leading to improvements in coverage. Additionally, knowing the coverage is especially poor in one or more supervision areas helps us choose priorities.

Principles to be applied to baseline studies:

1. Baseline data is always required.
2. Baseline studies can be time consuming and expensive. If possible existing secondary sources should be used to collect data.
3. Baseline study should be follow by an impact study which should use the same methods and study the same samples or sites to generate comparative data.
4. Baseline survey should be ideally done before the start of the project implementation.

General references about how to implement a survey are included in the Software WatSan Mission Assistant CD: Software Planning < Survey < How to conduct a survey (manuals):

- Strategies for Documenting Water & Sanitation Coverage and Numbers of Beneficiaries. Guidelines for quantitatively assessing and evaluating program requirements and accomplishments. IRC.
- Enquêtes CAP. ACF
- Knowledge, Practice and Coverage Survey. Field Guide. USAID.
- Sanitary surveying. WEDC Technical Briefs.
- Statistical procedures for nutritional surveys. MSF
- Using LQAS for baseline survey and regular monitoring. TALC.
- Guidelines for using mini-survey (OXFAM)

Phases of a Baseline Survey:

- **Phase 1:** Pre-implementation. This includes meeting with stakeholders and local experts, assessing data needs, developing and pre-testing a questionnaire, designing a sampling strategy and training supervisors and interviewers.
- **Phase 2:** Field Implementation Phase, which involves the actual collection of data in the selected communities.
- **Phase 3:** Post – Implementation Phase, which involves tabulating and analyzing the data, disseminating findings and using the data for decision-making.

U Carrying a survey requires certain level of knowledge and some specialist support is often required for key steps of the process (sampling strategy design and data analysis).

U Other agencies might also be doing surveys whose outcomes would complement our data (as nutritional surveys, health surveys, etc). It is recommended to coordinate and collaborate as much as possible with the other actors present in the targeted area.

Time allocation:

Activity	Time allocated	Resources required
Select key indicators based on program objectives and expected results. Determine what data are needed to measure the selected indicators.	Done at this stage (project logical frame)	Diagram of indicators
Developing a sampling strategy: - Work out sample size and sample unit - Decide on sampling procedure	One day	Stationary Computer - printer
Developing a questionnaire Prepare questionnaire and tally sheet Print questionnaires	One day	Printing costs for questionnaire Stationary Computer - printer
Recruit supervisors and interviewers	Two days – a week	Payment for advertisements Transport costs for interviewees
Training of supervisors and interviewers Translation into local language, Pre-test and review of the questionnaire Back translation into the original language to correct translation errors.	Two days	Training materials Transport for field visits Interpreter Per diems
Data collection Each questionnaire should take 30 minutes Each enumerator can do 6[1] questionnaires a day	Two or three days Depending on distances	10 enumerators for one day or five enumerators for two days Two days for contingency Transport Per diems for supervisors, drivers and interviewers Clipboards and pencils
Develop a data entry and analysis plan Data recording / Tabulating Data analysis Report writing and prepare the presentation	Four – Five days	Designated data recorders Tally sheets Final tally sheet The team leader The supervisors Pre-prepared tables Statistical analysis (appropriate computer software programmes).
Total time	15 days minimum	

[1] Please note that this is only a tentative indication since number of questionnaires depends on distances between households, comp tenancy of interviewers, availability of interviewees and the type of survey conducted.

A) Developing a questionnaire

Different examples of questionnaires are included in the Software WatSan Mission Assistant CD: Software Planning < Survey < Examples of Questionnaires:

- IFRC PHAST Baseline Survey.
- Kakuma Refugee Camp KAS Survey Template (ACF).
- KAP Public Health Questionnaire (OXFAM).
- Spatial Sanitary Survey Form (IRC).
- WatSan Household Survey (Nigerian RC)
- Ivory Coast Questionnaire Survey (ACH).
- Mini-Survey for baseline data (OXFAM).

Specific guides, questionnaires or checklists have been developed for conducting social research for hygiene communication campaigns. These in-depth interviews guides can be found in the Software WatSan Mission Assistant CD: Software Planning < Formative Research for HP campaigns:

- HP Background information – Checklist (EHP)
- Sample In-depth Interview Guide (EHP).

The specific questions listed in the questionnaires should be linked to the objective of the study (measuring water, sanitation and recommended hygiene behaviour) and must be justified in relation to the programme's indicators reflected in the narrative and log frame.

Adapting the questionnaire:

The questionnaire may be designed to gather basic data - for example, "Do you practice hand-washing at key times?" This type of questionnaire is a *quantitative questionnaire* as the results can be added up. But the questionnaire may also be designed to get more information in which case you do not want to ask a question which can be easily answered by "yes" or "no". In the example above the question asked might be "What do you think about creating a water and sanitation community committee in your village?" This is what is known as a *qualitative questionnaire*. Any question that is answered by 'yes' or 'not' should be followed up by other question that allow the respondents to explain or tell their stories. The best approach for qualitative questionnaires is starting with general questions about life situation and then moving on specific hygiene topics.

Regarding questionnaires it is important to be careful about structure of the questions to avoid biasing the responses, training the interviewers and building some cross-checking mechanisms to ensure consistency (for example you can include observation of practices at the household level or water point with information gained from more in-depth qualitative discussions with people). These will help you to ensure that questionnaire is effective in obtaining the information you need.

Pre-testing the questionnaire:

Using some participatory research will make the adaptation more precise and relevant to the local context. A pilot test of the questionnaire will show any issues of comprehension, relevance and reliability. E.g.: Specific model questions need to be adapted to the local context. For example, not all coding categories for water sources or toilet facilities may apply. Using some key informant interviews will make the adaptation more precise and relevant to the local context.

One model question can be asked in different ways. Test the questionnaire with a few community members in order to obtain feedback on their understanding of the wording of the questions. For example the question "what do you think about…" may be better structured by saying "what is people's opinion on…" in different contexts. The question 'what do usually you do when a child passes stools?' become closer to interview's true behaviour is you formulate the same questions as 'What did you do last time your child passed tools?'

Questionnaires need to be translated into the local language. This may be a source of error during the adaptation process. To minimize the risk of changing the meaning of the questions and the coding categories the questionnaire should be retranslated back into the original language - this will show the discrepancies that need to be addressed -.

B) Sampling Strategy

Sampling means collecting data from a group in the population that is representative of the whole. It has been likened to eating a bowl of rice where you only have to try one spoonful to know if the food is good enough to eat.

B. 1.- Population size and family size

Estimates of the total population represent the basis for all planning WatSan programs. These estimates may exist from a prior registration exercise or census, but often they are unreliable. The ideal method for estimating the total population size is by a census or registration system, which can only be carried out in months so for a rapid estimation the following steps:

B.1.1.- Mapping: When designing a survey it is important to take aspects of the local context into account since some of the following aspects should be explored in the survey:

- Geographical distribution of the population
- Religious, language and ethnic groups
- Castes / Tribes
- Different households structures (polygamous, female-headed, child-headed, etc)
- Socioeconomic groups.

It helps to have a visual image of the program area and the different factors that characterize the population so in case of not having maps available, consider preparing one. Begin with a tour around the boundary of their location to define the approximate shape, and the maximum and minimum length and width. If possible, the varying population density within the location should be shown.

B.1.2.- Determining population's size: Divide the area into sections using the previous map containing approximately the same number of households. To estimate the number of households in the entire location, count the number of households (houses, shelter or cooking fires) in a typical section and multiply this by the total number of sections. Then, carry out ***sampling***[2] and select a ***representative***[3] number of households. Record the number of persons living in each household, including theirs age and sex breakdown. Calculate the average number of persons per household and multiply this by the total number of households.

Sampling Unit

When dealing with large population groups it is not feasible to survey all individuals. However, valid conclusion can be drawn from measurements made on only a limited number of individuals within the population, provided that this sample is representative of the population as a whole. A project can collect information from a sample (a group of units – such individuals or households – selected from the general population) rather than from every person. Sampling units can be individuals, households or communities, depending of the focus of the study. If the sample is selected randomly, findings from the sample should generally reflect what is going on in the larger population.

Data gathered from a sample of a population provide only an estimate of what the results would be if measurements were made on the entire population. Whenever a sample is drawn, there is a risk that it may not be truly representative and therefore yield data that do not reflect the true situation, slightly different results are likely be obtained. A 95% confidence level (it means that the level represents and error risk of 5%, meaning that, out of 100 surveys, as many as 5 may give results that can not reflect the true situation) is usually considered to be appropriate for WatSan surveys. The precision of the result and the size of the confidence interval depend on the sample size and the actual prevalence of risky hygiene practices in the population.

---

[2] *See section on Sampling Procedure.*
[3] *See section on Sample Size*

B. 2.- Defining sample size

The sample size is the number of individuals to be included in the survey to represent each population of interest. The sample size required depends on the following factors:

- Required precision and confidence interval: the greater precision required, the larger the sample needed.
- Expected frequency of risky hygiene practices. The smaller the expected proportion of people presenting these risky practices, the greater the size of the sample required for a particular level of precision.
- Time and resources available: the time, equipment, transport and funds for the survey may limit the number of individuals or HH that can be visited.

The public health survey is meant to give an idea as to what is happening in the community – it is not a statistically correct study. However the sample must be large enough for you to comfortably assume that it is fairly representative of the majority of the population and small enough not to waste resources collecting from too many people. There are several ways of doing this:

B.2.1. Rough calculations based on population:

- For populations under 100 – 30-50 units[4]
- For populations between 100-300 use 50-70 units
- For populations between 300-1000 use 70-90 units
- For populations over 1000 use 90-100 units

Units might be households, groups or individuals.

B.2.2. Calculations based in *p* (expected frequency of risky hygiene practices among the population, as it is not known before the survey is done, an estimate must be used – this is an experienced guess, or derived from a small pilot survey) and *r* (relative precision required).
Example:
Expected frequency of risky hygiene practices 15% $p = 0.15$
Relative precision required 20% $r = 0.2$
$N = [(1.96^2 \times (1-p)] / [(p \times r^2)] = [(1.96^2 \times (1-0.15)] / [(0.15 \times 0.2^2)] = 544$

The sample size for cluster survey is likely to be larger than that for a random sample for the same precision. This is because the units within a cluster tend to be similar in their characteristics. When we calculate the sample size for cluster survey it is recommended to include a design factor (k = 2) in the formula.
$N = [2 \times (1.96^2 \times (1-p)] / [(p \times r^2)] = [2 \times (1.96^2 \times (1-0.15)] / [(0.15 \times 0.2^2)] = 1.088$

Projects that are interested in comparing changes over time (comparing baseline with a final survey) will need to collect data on control groups – communities that are not beneficiaries of the project activities but who are otherwise similar to the communities that are being targeted by the project).

B.2.3. Resources in Internet:
Determining sample size with simple software directly in the web site:

- http://www.isixsigma.com/offsite.asp?A=Fr&Url=http://www.surveyguy.com/SGcalc.htm
- http://www.surveysystem.com/sscalc.htm

EpiInfo can be downloaded from www.cdc.gov and has a section for finding sample size. A certain degree of familiarity with statistical terms is required.

B.3.- Sampling procedure

Commonly three main sampling methods are applicable to hygiene improvement – random, systematic and cluster. In both the random and systematic sampling a population list or register is required to

[4] Dale R. *Evaluation Frameworks for Development Programmes and Projects.* New Delhi: Sage; 1998

calculate the sample size, but in cases were the population size is not known and the population is in scattered clusters, cluster sampling is much preferred. This is also ideal in situations were there is limited time to train enumerators in random and systematic sampling methodologies.

B.3.1.- Simple Random Sampling:

The method we recommend here is the random sampling though an up-to-date list of all individuals in the population is needed with enough information to allow them to be located.

When randomly select units from the general population you ensure that every unit has an equal chance of being included in the study. Random sampling involves selecting units based upon chance. Simple Random Sampling (SRS) requires a ***sampling frame*** (a listing of every unit in the population, persons, households, villages, etc). With SRS every unit in the sampling frame is assigned a unique number. Then, a sample is drawn by randomly selecting numbers until you reach the desired sample size.

Any complete and up-to-date listing of all units in the total population can be used. The following are some examples: census, voter registration list, tax list, community health workers register, surveillance records and maps of the area showing each dwelling. When you have a complete and up-to-date sampling frame, you can use different methods as '*Random number table'*.

Example: You are given a list of 19 Communities and the Total Population for a fictitious supervision area.

a. Calculate the cumulative population of 19 communities: Begin by adding the population of the second community (730) to that of the first (548) and writing the total (548 + 730 = 1.278) in the first blank space in the far right column, 'Cumulative Population'. Repeat this process by adding the population of the third community (686) to that of the combined population (1.278) to get the new total (686 + 1.278 = 1.964). Write it in the blank space in the far right column. Then do the same for the next community adding its population (280) to the previous total (1964) to get the new total: 280 + 1964 = 2.244. Fill in the 10 remaining blank lines at the bottom of the chart. The total of the population sums up 23.489.
b. Calculate the Sampling Interval. The answer is 23489/19 = 1.236,26.
c. In this particular instance we are using a random number to help us identify interview locations. A Random Number Table that has 14 columns made up of rows of random numbers. You can use any randomizing process you wish, but using a random number table is recommended.
d. Restate the number of the sampling interval (1.236,26) fixed in the previous step. The random number has to be between 1 and the sampling interval, 1236 (the decimal point is not used in this step.)
e. Identify the highest possible number of digits in the random number, which in this case is 4, the number of digits in the interval (1…2…3…6). Decide which of the five displayed on the table they will use in this particular case (it is recommended to use the first four).
f. Close their eyes and hold a pencil in the air over the random number table. Bring the pencil down on the table while keeping your eyes closed. The pencil should strike on or near a row of random numbers near one of the columns of numbers. Using the first four digits, is the number in the range of 1 and 1236? If it is not, move to the next row and keep doing this until they find a 4-digit number in this range. *That* number is a random number that could be used in this example. Let's assume the random number selected is 0622. The location number of the 1st interview is the random number. For this demonstration, we are assuming that random number 622 was selected in the previous step. The location number of the second interview is equal to the random number plus the sampling interval, in this case 622 + 1.236,26= 1.858,26 (for this step you always use the decimal).

If a 'random table' is not available you can assign each household on the list an identification number. A number corresponding to each household is written on a small piece of paper, which is placed on a large box. The pieces of paper are shuffled and picked out blindly. The households selected in this way become the sample for the survey (they can not be excluded or substituted for any reason). If you do in public the community can see how the households are selected.

B.3.2.- Systematic Sampling:

Systematic sampling eliminates the need for complete, up-to-date population registers, but requires:

- A reasonably accurate plan or map showing all the households, and
- An orderly layout, or site plan, which makes it possible to go systematically through the whole site.

This technique has been used in well-organized villages where households are arranged in blocks and lines. The procedure is as follow:

- Trace a continuous route in the map, which passes in front of all the households.
- Calculate the number of households that must be visited (*Sample Size*[5]*)*.
- Calculate the sampling interval by dividing the total number of households by the number that must be visited. If the total number of households is 5.000, and 363 are to be visited, the sampling interval is 5.000/363 = 13.8, or 13 (round down to the nearest whole number in this calculation).
- Select the first household to be visited within the first sampling interval at the beginning of the route by drawing a random number (e.g. random number table) which is smaller than the sampling interval (e.g if the number drawn is 7 start with the seventh house).
- Select the next house by adding the sampling interval to the first selected house along the prescribed route.
- Continue in this way until the number of households required for the survey has been systematically selected (they can not be excluded or substituted for any reason).

B.3.3.- Cluster Sampling:

Other methods for SRS – Simple Random sampling -exist and they are valid. Especially if communities are far apart or very big, *cluster sampling* is a good approach, as logistically it is easier. Give all communities (or sections in a camp) a number and then select about 5% of these by picking numbers out of a hat or using random tables. If communities are very different form each other (for example some are in the hills and some are by the sea) then you need to make sure you select from both. In the case of cluster sampling, the sample size should be 50% larger than when using simple random selection.

What is a cluster? A cluster is a naturally occurring group of individuals (such as a village, ward, or city block – when natural groupings do not exist, artificial clusters may be defined by imposing a grid on a map of the area) likely to include the population group your project is interested in studying. Cluster sampling is a very popular method due to:

- It does not require a sampling frame, other than a list of population centers (such as towns, villages or communities), their estimated population size and their accumulative population. Where feasible, the population is divided into a large number of clusters containing similar number of people (well defined villages of similar size are examples of possible clusters, larger villages can be divided in two or more clusters.
- By interviewing a number of people who live in the same cluster, it reduces time and travel costs between interviews.

[5] *See section Defining Sample Size*

Sampling is done in two stages:

i) Cluster sampling[6]:

a. Calculate the sample size based on the desired level of precision and confidence.
b. Determine the number of interviews per cluster (it is suggested to conduct 10 interviews in each cluster).
c. Divide the sample size by the number of interviews in each cluster. This will give you the number of clusters.
d. Prepare a list of all existing units with their estimated populations. Add two more columns. In the first, record the cumulative population figures obtained by adding the population of each unit or zone to the combined population of all the preceding units or zones on the list. Note: the cumulative population of the last community listed in your sampling frame should equal the total population of the entire program. If this is not the case, re-check your calculations.
e. Calculate the sampling interval by dividing the total population of the entire program area by the total number of clusters required.
f. Choose a random number. This number will be used to identify the starting point on the list to begin selecting clusters. The random number must be less than or equal to the sampling interval.
g. Look at the column where you have listed the cumulative population of each community and determine which community contains (that is, the cumulative population equals or exceeds) the random number.
h. To identify the second community where a cluster is located, add the sampling interval to the random number selected in step f. The community whose cumulative population equals or exceeds that number is the location of cluster 2.
i. To identify the remaining clusters, add the sampling interval to the number that identified the location of the previous cluster.

Example:

- Sample size: 300
- Number of clusters. If you plan a sample size of 300 and doing 10 interviews in each cluster, you will have 30 clusters in your survey.
- Total population in the program area: 301.170
- Sampling interval = 301.170 / 30 = 10.039 (You can round the number to the nearest whole number, e.g. 10.040)
- List of units and cumulative population

No	Name of community	Population	Cumulative population	Cluster
1	Utaral	12.8888	12.888	1
2	Bolama	3.489	16.377	
3	Talum	6.826	23.203	2
4	Wara-Yali	4.339	27.542	
5	Galey	2.203	29.745	
6	Tarum	4.341	34.086	3
7	Hamtato	1.544	35.630	
8	Nayjaff	885	36.515	
9	Nuviya	2.962	39.477	
10	Cattical	4.234	43.711	4

[6] A common cluster sample survey includes 210 households per district – 30 clusters with seven households per cluster.

11	Paralal	1.520	45.231	
12	Egala-Kuru	3.767	48.998	
13	Uwanarpol	3.053	52.051	5
14	Ilandia	60.000	112.051	6-7-8-9-10-11
15	Puratna	2.207	114.348	
16	Kagaini	1.355	115.703	
….				
50	Yandot	3.193	139.309	30

- If the sampling interval is 10.039 you would select a random number between 1 and 10.039. As an example it might be 9.679.
- Look at the column where you have listed the cumulative population of each community and determine which community contains the random number. Utaral, the first community listed in the sampling frame has a cumulative population that equals or exceeds the random number so this is cluster 1.
- To identify the second community where a cluster is located, add the sampling interval (10.039) to the random number selected (9.679). The cluster 2 is in Talum because 10.039 + 9.679 = 19.718 and the cumulative population in Talum include this number. The cluster 3 is in Tarum because 19.718 + 10.039 = 29.757 and the next community with cumulative population including this number is Tarum. Cluster 4 is Cattical since 29.757 + 10.039 = 39.796.

ii) Selecting individuals within each cluster.

- A sketch map of the area might be drawn, the houses numbered and the household selected using a random table.
- '*Spin the bottle'* option is recommended if the site layout does not permit the previous option. This technique is used to identify the starting point within a sample area. Spinning a bottle at the center of the unit or cluster – usually the point where the population is about equally distributed on all sides - helps the survey team to randomly choose a direction to follow. Walk in that direction from the centre to the outer perimeter of the unit or cluster, counting the number of households along this line. Visit all the households along the randomly chosen line, choosing the households with doors nearest to the last house surveyed.

Also there are alternatives *parallel sampling, stratified sampling*, etc, so we strongly recommend exploring the different methods in the existing bibliography and selecting the option than better suits with your project sampling need (general references about how to implement a survey are included in the Software WatSan Mission Assistant CD: Software Planning < Survey < How to conduct a survey (manuals)).

Examples:

- Random selection of samples (using Microsoft Excel):
http://www.isixsigma.com/library/content/t000702.asp

C) Training supervisors and interviewers

When considering who to use as an interviewer, it is best to consider skills such as communication and the ability to get on with the community. Confidentiality may be important and people can be assured that individual comments will remain anonymous. They need to be able to speak the local language and it is good to have a gender mix. Sometimes young women are not able to talk to older men or men cannot interview women so this needs to be considered when recruiting.

It is important to assess existing skills and identify areas where training and supervision should concentrate. The trainer should find out what members of the study team have already done and build on their knowledge and skills as much as possible, rather than introduce a completely new set of skills which will require more time to master. On the practical side, the focus should be on activities that enhance the team's investigative and analytical skills, such as:

Observation skills: Observers need to learn how to write systematic detailed descriptions of what is observed; and separate relevant detail from trivia without being overwhelmed by the amount of trivia. It is important to note that even skilled and well-trained observers can bring in their own biases to the information. Trainees should be encouraged to: discuss whether observers should let the observed know what they are doing or keep it to themselves, and why; suggest what the duration of the observations should be, and why; and discuss whether the focus of the observations will be broad or narrow, and why.

Interviewing skills: The two main skills required for successful interviewing are an ability to establish rapport with the interviewee and keen listening. There are several exercises, role plays and games that can be used to improve interviewing techniques, particularly in the areas of probing, listening, observing non-verbal cues and recalling the content of interviews.

Group discussion moderating/facilitating skills: Decisions about what methods/tools to try out during the training and which ones to prepare and use in the actual study should be based upon the study objectives, the capacity of the study team, and availability of time and material resources. Some techniques which can be used in group discussions are: community mapping, history line, seasonal calendars, gender roles, and focus groups. Some techniques for group dynamics can be found in the section on Group Dynamics in Research Techniques.

Never assume that everyone knows how to ask questions using a questionnaire. It is worth taking the time to do at least a days training on questionnaires, asking questions, not using positive or negative body language. It is also good to use the time to get the questionnaire translated into the local language (the enumerators can do this) and to pre-test it. This gives everyone a chance to familiarise themselves with the tool.

Chose one person to be the team leader for each group and make sure they understand the importance of checking questionnaires in the evening for mistakes. When the interviewers return every day with completed questionnaires, this person should take time to run through them and check for mistakes and “missing values” (blanks where there should be answers). If one enumerator is consistently forgetting to ask questions or makes mistakes, she/he should be able to pick it up early. If possible, enter the data onto the tally sheets every evening. Use the five-line method and then add totals.

D) Collection of data

For surveys the main method of data collection is ***household questionnaire***. A general questionnaire might include quantitative and qualitative data but also a specific quantitative questionnaire might be developed after gathering some information from the community through focus discussion groups.

Also we might develop some cross-checking mechanisms to ensure consistency and ensure that the information obtained is what you need as ***transect walks*** and ***direct observation***. Through these methods qualitative data might be collected as not only cross-checking information but also as complementary information. Lets interviewers check some household characteristics that are conditions for or the result of a hygiene behaviour, e.g.: it is much accurate to observe that soap is available to the household than rely on an answer only.

In communities where there already has been substantial hygiene education and they are aware what they should be following, respondents may simply try to please the interviewer by saying what they believe is the desired (correct) answer, regardless of what do or believe. Therefore, it is important that responses concerning current hygiene practices be corroborated as much as possible by observation.

Moreover questionnaires are useful to get very specific information, but are not useful to create a dialogue themselves. To generate dialogue, a suggestion would be to hold ***focus group discussions*** after the results have been analysed to share them with the community members.

E) Tabulating and analysing data

E.1 Tabulating

Tabulating is bringing together the information that was collected through the interviews in a form so you can analyse it. A system of collating the information will need to be devised, such as the use of *tally sheets*, where a line is used to record answers to each question and then grouped together in fives for easy counting.

The best way to analyze the information is by coding the answers (e.g. 0 = incorrect answer; 1 = correct answer; s = skipped question; x = missing response) and producing a summary tabulation table which includes all the codes resulted from the tabulating process. Then you can calculate the average coverage (percentage of people in a given area who know of and/or practice a recommended health behaviour or receive a particular service) to know if that indicator is below or above the minimum standard.

E.2. Analyze the questionnaire.

Go through all the sheets and make sure they are clearly filled in and completed. Look at how everyone answered each question. If it is a yes/no answer, a statistical figure can be extracted. For example, 56 of 80 people answered that "yes" they practice hand-washing, or 70% of those asked. Advance analysis of the quantitative data is frequently performed by experts using statistical software packages (as EpiInfo).

If the questions are not direct yes/no answers, try to get an overview of the ideas expressed and group similar answers together. For example, if the questionnaire looks at the causes of health problems in the community and most people answer that there is a lack of clean water and lack of education about sanitation, these answers can be obviously grouped together. Each questionnaire can be coded according to the issues raised by the participants.

In the CD an example of excel sheet used by Liberian RC containing a questionnaire form and a tally sheet linked to a graphics sheet can be found: Software Planning < Survey < Tabulating and analyzing data.

F) Reporting

Different examples of survey reports can be found in Software WatSan Mission Assistant CD: Software Planning < Survey < Examples of Survey Report.

- Diarrhea & Survey Nepal. UNICEF
- Environmental health Assessment. EHP.
- Kibondo Environmental Assessment. IRC.
- Kenema District Survey Report. OXFAM.
- Angola KAP Survey Report. ACH
- Cote d'Ivoire rapport d'enquête CAP (ACF).
- Guatemala Informe de Línea de Base (ACH).
- Laos rapport d'enquête CAP (ACF).
- Philippines KAP Survey Report. ACH

G) IFRC PHAST Base line Survey. Lessons learnt

The PHAST baseline survey questionnaire is not necessarily intended to be used uniformly by all Red Cross/Red Crescent National Societies but rather to be used as a template and adapted accordingly.

- The questionnaire is aimed at mothers/female caretakers.
- It should take approximately 30 - 40 minutes to complete one questionnaire
- The questionnaire will be completed by a volunteer. Often better for volunteers to work in teams of two especially if they are women.
- The questionnaire will be analysed at Red Cross/Red Crescent branch and/or headquarters level
- It should take at least 2 days to train a volunteer to complete the questionnaire.
- Where possible the volunteers who will undertake to complete the questionnaire have undergone PHAST or CBFA training.
- As it is rarely possible to include the entire target group in a survey, a limited number of people should be therefore chosen known as the sample. The characteristics of the sample should be similar to the total population so it is as representative as possible. The different ways to obtain a sample include simple random sampling (picking names from a hat or at random from a list, interval sampling (by selecting persons from a list at regular intervals or cluster sampling (where groups of people rather than individuals are selected to comprise the sample).
- The base-line survey should not take longer than 1-2 weeks to undertake. An example to calculate the time needed to undertake a base-line survey is to take a sample size of for example 100 households and hence 100 questionnaires to be completed. Suppose in one day a person or a team of two can complete 6 questionnaires, therefore it will take 1 person 17 days to complete the 100 questionnaires or 2 people 8.5 days (Note: you need to allow time for travelling to the community, and walking between houses and introductions).
- The base-line survey should not be followed up more than once a year.