



Problems in representative sampling in the water and sanitation sector: a brief analysis of problems and possible solutions

Author: Caroline Hunt, April 2006

Quality assurance: Kristof Bostoen

Edited and produced as a PDF document: May 2020

What is sampling?

Sampling is part of statistics, the scientific discipline and tool used for investigation of biological and medical science in particular. Sampling is the tool used to select part of a population for data collection and analysis. This selection, the sample, is then used as a manageable number of people or objects (depending on what is being investigated) to then form the basis for analysis. In many cases, collecting data for the entire target population would be too expensive in terms of time and resources, as well as too challenging logistically.

We use sampling in every day life. For example, if buying a large quantity of fruit or vegetables from a market, we would probably not check the quality of every item, but instead might check a few for damage and disease. From that we would then make a decision about whether we were happy with the quality of them all.

Sampling techniques can be borrowed and applied by many people in their work. For instance, the accuracy and credibility of field surveys in the water supply and sanitation sector can be enhanced by following accepted sampling techniques.

Why would you use sampling in the water and sanitation sector?

Sampling is used in the sector for a number of different reasons.

1. Monitoring

Monitoring in the sector is carried out by a large number of different actors and for a variety of reasons. At national level, large-scale household surveys are carried out to provide data representative of the whole nation. For instance, OCR-Macro carries out the Demographic and Health Survey (DHS) and UNICEF is responsible for the Multiple Indicator Cluster Survey (MICS). These national data sets provide information about both water supply and sanitation coverage which is used by the WHO/UNICEF Joint Monitoring Programme in monitoring world progress towards the Millennium Development Goals.

At more local level, numerous non-governmental organisations use sampling techniques in carrying out field surveys to collect information about their target populations, such as identifying their needs. This may involve field staff selecting households from the whole community and using a questionnaire to collect information from each of the selected households. Similarly, aid agencies might use field surveys to inform and evaluate their water supply and sanitation programmes.

2. Research

Numerous research studies use statistical sampling techniques in order to investigate a broad range of subjects concerning water supply, sanitation and hygiene. These include epidemiological studies which look at the health risks associated with water supply, sanitation and hygiene. Economic research work is another area and has previously included studies on willingness-to-pay for water supply and sanitation facilities. Further research work in the sector has covered a wide array of subjects such as water quality, institutional capacity and sustainability of behaviour change.

What is representative sampling?

A sample that is fully representative of the population from which it is drawn is called a *representative* sample. The sample needs to be representative in order to infer the results from the sample back to the whole population. Statistical analysis can only be used on representative samples, otherwise nothing can be said about the total population.

There are four main steps to enable inference from a representative sample.

1. Clearly defining the target population from which the sample is to be selected (for example, the population of a certain geographical district).
2. Clearly defining the basic sampling unit (such as the household).
3. Ensuring that each sampling unit has an equal or known chance of being selected into the sample (such as using random sampling where every sampling unit has been identified (such as households from a census) and every one has an equal chance of being randomly selected).
4. Take in to account the sample design in the analysis of the data.

In the water and sanitation sector there are a number of issues, mentioned below, which make carrying out these steps problematic.

What are the problems in using representative sampling in the water and sanitation sector?

1. Clearly defining the target population from which the sample is to be selected

Firstly, defining the whole of the target population is often difficult when information is lacking about households. This is a problem particularly for informal and low-income settlements where such information is usually lacking. For instance, if a local authority is providing a list of addresses for a certain geographical area, it is unlikely to include addresses for those within informal areas. The sample subsequently selected is therefore likely to be biased towards more affluent households, and so not fully representative of the whole population. Because sources of bias can have a strong detrimental impact on the results of data analysis, it is always important to fully document the sampling process and to follow up on non-responding households to make sure that there is not in effect a distinct sub-sample missing from the analysis. This might be an issue if all households were visited in the daytime and some of the households were empty. Empty households may represent a different group if their members are out working (they may have different household income levels to those who do not work).

2. Clearly defining the basic sampling unit

For much of the sampling undertaken in the water and sanitation sector, the basic sampling unit is the household. This is because all of the members of the household are likely to use the same water source (such as a piped water supply or a well). This is perhaps slightly less true for sanitation, where some household members, such as children, may not use the same sanitation facility as other household members. When a survey requires information on the state of public water sources (such as the demand on these sources) then the household might not be the sampling unit of choice. Instead, the water source would be the basic sampling unit.

3. Ensuring that each sampling unit has an equal or known chance of being selected into the sample

This third step of taking a representative sample can be extremely difficult because of lack of information during the first step (the population sampling frame not being available). If all of the sampling units can be identified (for example, all addresses are known), it is possible to use simple random sampling. In most instances this is not the case and so different sampling techniques have to be used instead. Any form of randomisation during the sampling process will reduce the risk of sampling bias. Some alternatives to simple random sampling are mentioned in the paragraphs below.

What are the possible solutions?

For many people having difficulties, seeking advice can be the best starting point. In research studies, statisticians are usually consulted early on when the study design is being determined. Statisticians can help with advice on sample sizes and sampling techniques.

Some potential solutions can come through using good practice in carrying out surveys. For instance, having a pilot study (a small trial survey before the main survey) can provide a good understanding of the difficulties involved with the work to be undertaken and can help in problem solving later on. Secondly, good training of field staff can enable greater understanding of the process and why and how to avoid mistakes on the ground such as using convenience to choose households rather than sticking to the sampling protocol. Thirdly, thorough documentation of the sampling process can help enormously in the analysis and writing up stages, as it provides a record of what has been done, and how.

Large-scale sample survey providers use techniques such as cluster sampling, stratified sampling and weighting. These techniques can also be used to the advantage of different actors within the water supply and sanitation sector to attain greater accuracy and cost-efficiency in their sampling.

Cluster sampling reduces the need for population lists for the selected cluster. Because people usually live in clusters (such as towns or villages), these clusters can be used as the basis for sampling. For instance, when using clusters of villages, lists of addresses can then be developed for each of the selected villages. These lists are then used to draw samples from (for example, by random selection through using random numbers tables or by pulling addresses out of a hat). Multi-stage cluster sampling would then involve taking a next set of samples from these lists (for instance, listing all households with children under five years to form a further sample). Cluster sampling can be a very efficient and cost-effective way of carrying out sample surveys.

Stratifying is where the target population is divided into different groups ('strata'). Samples are then selected from each strata. It is a technique used when comparisons are needed between different groups, as well as requiring estimates about the total population. An example would be if detailed

information were required about different social or ethnic groups or urban and rural populations. To carry out stratified sampling, sufficient information is usually required prior to the data collection.

Most common sampling methods will aim to have a self-weighted sample. This means that no sample weights are required. Weighting of data is used to correct imbalances in the probabilities of selection. For instance, if village A has twice as many respondents in the sample than village B (but A has a much smaller population size than B) then weighting can be used to weight up the under-represented village B respondents and weight down the over-represented village A respondents. Weighting is important as it has an impact on the results but is best done by statisticians using appropriate software.

As seen by the challenges mentioned above, the water and sanitation sector is in need of more suitable sampling methods. These methods will need to be able to deal with situations in which there is a high degree of clustering (often making the construction of sample frames impossible (Bostoen, in print)). These methods could also be designed to be more user-friendly for people working in the sector (who are often unfamiliar with survey statistics).

Bibliography

Bostoen, K. and Chalabi, Z. (in print). "Optimising Household Survey Sampling without Sample Frames." *International Journal of Epidemiology*. <http://ije.oxfordjournals.org/cgi/reprint/dyl019?ikey=JNguy1mhrzW6D3u&keytype=ref>

Additional sources of information

WELL monitoring factsheet

<http://www.lboro.ac.uk/well/resources/fact-sheets/fact-sheets-htm/mwsc.htm>

Guides on how to sample

Kalsbeek, W and Heiss, G. (2000). "Building Bridges Between Populations and Samples in Epidemiological Studies," *Annual Review of Public Health*, 21:1-23.

http://www.sph.unc.edu/chsr/Dissemination/Arph_sub.htm

Hunt N & Tyrrell S. (2001) DISCUSS Sampling Methods. Coventry University

<http://www.mis.coventry.ac.uk/%7Enhunt/meths/listof.htm>

FANTA Publications Sampling Guide. Food and Nutrition Technical Assistance. USA.

<http://www.fantaproject.org/publications/sampling.shtml>

Grosh, M. & Glewwe, P. A Guide to Living Standards Surveys and Their Data Sets. LSMS Working

Paper Number 120, The World Bank, 1995 <http://www.worldbank.org/LSMS/guide/describe.html>

Glossary on sampling

[Statistics Glossary - sampling](#)

Looking for alternative sampling methods

Bostoen K, Chalabi Z. Optimization of household survey sampling without sample frames.

International Journal of Epidemiology 2006 Feb 15.

<http://ije.oxfordjournals.org/cgi/reprint/dyl019?ikey=JNguy1mhrzW6D3u&keytype=ref>

Getting statistical support on sampling

Using the WELL Technical Enquiry Service if you work for an eligible organisation

WELL - Resource Centre Network for Water, Sanitation and Environmental Health

<https://www.lboro.ac.uk/research/wedc/rc/types/km/well/>