Measuring access and practice

(Water Sanitation and Hygiene)

Upgrading,30 October 2002

Kristof Bostoen

Supervisor:	Upgrading panel	
Professor Sandy Cairncross	Internal:	Dr. Steve Bennett Dr. Adam Biran Dr. Ursula Blumenthal
	External:	Kathleen Shordt

2025-17511

Table of contents

List of Tables

List of Figures

List of Equations

Acronyms, Abbreviation and Glossary

ArcGIS™	commercial GIS software product line by ESRI™
BSU	Basic sampling unit, the unit from or for which the data is collected
CMD	Camp Dresser & McKee International Inc. Consultants in USAID's EHP programme
Cluster	Selected primary sample unit (PSU)
DALY	Disability Adjusted Live Years is an indicator based on 'time lived with a disability' and 'time lost due to premature mortality', developed for the World Bank (World Bank 1993)
DHS	Demographic Health survey by MACRO for USAID
EHG	Environmental health group at the LSHTM
EHP	USAID's Environmental Health Project
EPI	WHO Expanded Program of Immunisation. Also a *sampling method* used in this program.
EPSEM	Equal probability of selection method
ESRI™	Environmental Systems Research International™ company developing ArcGIS™ products
GA2000	Global Water Supply and Sanitation Assessment 2000 (WHO et al. 2000)
GIS	Graphical Information System
HH	Households
IAP	Iguaçu Action Plan by WSSCC for the water, sanitation and environmental hygiene sector
JMP	Joint Monitoring Programme of UNICEF and WHO
LSHTM	London School of Hygiene and Tropical Medicine
Macro Inc.	Macro Int. Inc. part of ORC Macro and responsible for DHS surveys for USAID
MICS	Multi Indicator Cluster Sampling
Population	In statistics population means all the BSUs that have a non-zero-chance of being selected; It is also the level at which statistical inference is made.
PPS	Probability Proportionate to Size.
PPES	Probability Proportionate to Estimated Size.
PSU	Primary sample unit
Self weighted	Equal non-zero probability (see also EPSEM)
SRS	Simple random sample
Take	Take is the amount of samples (sample size) taken in a cluster.
UNICEF	United Nations International Children's Emergency Fund
USAID	United States Agency for International Development
V21	Vision 21, agenda for the water, sanitation and environmental hygiene sector by WSSCC
VBA™	Visual Basic for Applications ™ Microsoft ™ script language
WHO	United Nations World Health Organisation
WSSCC	The Water Supply and Sanitation Collaborative Council has a mandate of the UN general assembly to coordinate activities that accelerate the achievement of sustainable water, sanitation and waste management services to all people, with special attention to the unserved poor.

1 Introduction

Access to safe water and sanitary means of excreta disposal are a universal need and should be considered basic human rights. They are essential elements of human development and poverty alleviation and constitute an indispensable component of preventive healthcare. With Vision 21, the Water Supply and Sanitation Collaborative Council (WSSCC) has set the agenda for the water, sanitation and environmental hygiene sector in the first quarter of the 21st century. Parts of the Vision 21 agenda have been adopted by the World Summit on Sustainable Development in Johannesburg[1]. Vision 21 set clear sector targets for the year 2025. Out of the six targets identified by Vision 21 (Table 8, Annex A), the three first include 'reduction of people lacking safe water and access to adequate sanitation by 50%' as well as 'universal public awareness of hygiene' by the year 2015. In order to attain these goals it is important to measure and quantify access to water and sanitation as well as measure good hygiene practices. To achieve this the 'Water Supply and Sanitation Collaborative Council' (WSSCC) indicated in its 'Iguaçu Action Programme' (IAP) the need for better universal measuring methodologies for measuring access to water and sanitation.(WSSCC 2002) Some of the Collaborative Council's main action points on monitoring are to define, test and validate a core set of indicators for measuring Vision 21 implementation as well as to build consensus on methodologies for data collection which can ensure a uniform standard for analysis and reporting of progress towards these goals.

2 Objective of the research

The objective of the research is to design a survey methodology to measure access to 'improved' water and 'improved' sanitation as well as adherence to 'improved' hygiene practices, with a focus towards the Vision 21 targets (WSSCC 2000b) as based on the Iguaçu action programme (WSSCC 2000a).

The specific research question is:

Can we measure in a *representative* way, and at *reasonable cost, summative information* on the population's:

- adherence to 'improved' hygiene practices;
- access to 'improved' sanitation;
- access to 'improved' water sources,

in a *specified region* such as a country, province or district?

Table 1: Research question

2.1 Targets of Vision 21

Vision 21 has 6 targets as shown in Table 8 (Annex A) of which the three included in this research are briefly discussed below. Their achievement is planned by the year 2025. Intermediate targets have been set out for the year 2015 as shown. Those targets that have been adopted by the Johannesburg world summit have been marked with footnotes.

Target 1

	Vision suggested targets for 2015	for 2025
1	Universal public awareness of hygiene	Good hygiene practices universally applied

The above are abstract ideals rather than clear targets. It would be better to measure the same 'application of good hygiene practices' or 'awareness of hygiene' during both periods to enable comparison. As it is practice rather than awareness that is an engine for change and also to obtain a more quantitative objective for 2015, the following is suggested as a working definition:

1	*Halve the percentage of people not applying good hygiene practices.*	Good hygiene practices universally applied

To measure progress toward this intermediate target there is a need for baseline data which is not yet available. Another problem in achieving these targets is the definition of 'good hygiene practices'. For the purpose of Vision 21 we would suggest the following definition for good hygiene practice:

Day to day application of practices and habits reducing risk of faecal-oral transmission of pathogens.

This definition focuses on faecal-oral transmission as the biggest cause of hygiene-related morbidity and mortality (WHO 1992) largely preventable through access to water, adequate sanitation and good hygiene practices. Measuring '*universal application*' of good hygiene practices is difficult to do, in particular if it means 100% application. For that reason we suggest that for working purposes "universal" is defined as *90% (±10% points)* of households applying good hygiene practice, which would make it more useful as a statistical reference.

Target 2

	Vision suggested targets for 2015	for 2025
2	Percentage of people who lack *adequate* sanitation halved[1]	*Adequate* sanitation for everyone

In its 'Global water supply and sanitation assessment 2000', the WHO / UNICEF joint monitoring programme (JMP) no longer refers to '*safe*' drinking water and '*adequate*' sanitation in their reporting. Instead, '*improved*' water and sanitation technology types are now used as the criterion of coverage. This change in terminology reflects both past misrepresentation, and future uncertainty, in judging and defining services as 'safe' in terms of human health (Hunt 2001). For that reason it is suggested that the same terminology be used in the interpretation of Vision 21 targets. It is also the terminology used in the rest of this document.

2	Percentage of people who lack *improved* sanitation halved	*Improved* sanitation for everyone

Sanitation here is seen in its narrowest definition as 'human excreta disposal' and will be regarded as such in the rest of the document.

If access '*for everyone*' means 100% access than this will be difficult to measure and to achieve as mentioned above. For statistical reasons we will define 'everyone' as *90% (±10% points)* of households having access to improved sanitation.

[1] The Johannesburg World Summit adapted the goal of halving, by the year 2015, the proportion of people who do not have access to basic sanitation

Target 3

	Vision suggested targets for 2015	for 2025
3	Percentage of people who lack *safe* water halved[1]	*Safe* water for everyone

Here as well the JMP terminology of '*improved* instead of '*safe*' is suggested for the same reasons as mentioned before. Reformulating the targets would be as follows:

3	Percentage of people who lack *improved* water halved	*Improved* water for everyone

Discussion of what '*improved'* means is still an issue. In this document 'Having improved water' means:

- Use of drinking water which is protected from faecal pollution.
- Availability of enough non-drinking water for basic hygiene purposes.

To enable progress to be measured on these targets, baseline information must be available. Some of this baseline information is available in the form of the 'Global Water Supply and Sanitation Assessment Report Year 2000' (WHO et al. 2000) and comparability with existing data should be maintained where possible.

2.2 Planned survey methodology

The monitoring or surveying for which the JMP was established (with a mandate from the UN Secretary General), and which the WSSCC was mandated at Iguaçu to promote, is principally "summative". Its aim is to measure a small number of quantitative indicators to determine whether targets are being achieved (Cairncross 2001). Summative information is only concerned with characterisation of a situation while formative information is more analytic, seeking a diagnosis of problems needing resolutions.

The household is used as Basic Sampling Unit (BSU), so the outcome of the survey is a dichotomous value by household for each target. These will indicate for example, whether the household has or does not have access to water. This means that all data we assume to be important and relevant for each target have to be combined until they reach a *'yes'* or a *'no'* value for that particular indicator.

Targets 1-3 require outcomes for each household that can lead to the conclusion that the household has:

1 *Good hygiene practices Yes/No,* meaning that the behaviour of the household is such that it reduces the risk of pathogenic transmission.

2 *Access to adequate sanitation Yes/No,* meaning that excreta is disposed of in such a way that it reduces the risk of faecal-oral transmission to its users and the environment.

3 *Access to improved water supply Yes/No,* meaning that they have access to sufficient drinking water of acceptable quality as well as sufficient quantity of water for hygiene purposes.

2.3 Brief outline of the project

The project is to design a simple cross sectional method to measure Vision 21 targets. The project can be spilt into three different parts of which only the first two are part of the research project.

[1] The Johannesburg World Summit reaffirmed the 'Millennium Development Goal' of halving, by the year 2015, the proportion of people without access to safe drinking water.

The third part has to be kept in mind when designing the first two. A detailed project plan can be found in Annex C.

Figure 1: Major parts in the project

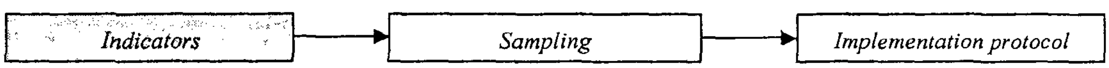
The first part of the project is identifying representative data that can be easily and timely collected in a cross sectional survey at a low cost and with local means. The data collection has to be as universal as possible so it can be used in various contexts and cultures without modification. The relation between the data and the indicators has to be transparent.

The second step is to develop a sampling method without sampling frame, similar to the method developed for the WHO/UNICEF's Expanded Program of Immunisation (Levy et al. 1999). This should be a simple statistically valid method enabling statistically un-trained people to collect data in an economical way and allow for simple analyses of the data collected.

There should also be a clear protocol for situations where sample frames are available.

For this method to be useful it must be accepted by the international water and sanitation community as authoritative. There will also be a need to design clear protocols and guidelines allowing people to adhere to the conditions and the methodologies defined in the first two steps. The political process of ensuring acceptance and the drafting and dissemination of guidelines is however not part of the research.

3 Indicators



The assumption behind the survey is that health in the community can be improved through access to water and sanitation as well as the application of hygiene practices. The method wants to measure who has 'access to', or 'practise what' is considered beneficial for their well being. However these values are relatively abstract and difficult to measure. This can be partially solved by developing indicators.

To take an analogy from bacteriology, it is practically impossible to measure pathogenic concentrations of the cholera vibrio, Salmonella typhi, and every other currently known disease-causing organism in drinking water or in food on a regular basis. It has, however, become relatively straightforward to measure the indicator organism *E. coli*, and this indicator has been widely successful in advancing the objective of "safe drinking water" to promote the value of "Health for All." (Kolsky et al. 2002).

Performance indicators are defined in this paper as practically useful surrogates for the direct measurement of performance. Most standards are based on indicators, because they can be measured reasonably easily, rather than the performance itself.

Indicators are by definition, "an indication" of status or process rather than the measurement of the status or process itself. Indicators are inherently open to debate, precisely because they are imperfect surrogates for what they 'indicate'. The question continually asked during the whole process of choosing indicators is whether the indicator reflects accurately enough the critical aspect of the performance.

3.1 Data collection

In public health the use of sophisticated research designs and statistical techniques are only as strong as the data used for these analyses. The quality of the data depends on the quality of the sampling and measures used in their collection. The collection of accurate data is the foundation for all scientific data analysis.

There are many ways of systematic data collection. For hygiene related issues, summative information is generally collected through questionnaires, observations and demonstrations. Systematic data collection by using **questionnaires** has some powerful advantages over less structured approaches, as well as some limitations (Curtis et al. 1993; Pedersen 1994) as shown in Table 2.

Advantages:
• Efficiency: simple and cheap to administer;
• Consistency, comparability, generalisation: Standardised formats ensure all respondents are asked the same question
• Summary and analysis: it provides quantitative data that can be quickly summarised.
• Scientific rigour: questionnaires can evaluated for reliability, validity and responsiveness
Limitations:
• Limited depth: cannot generally provide an in-depth view;
• Inflexibility: structured, standardised format is constraining;
• Can't detect the unexpected
• Error and bias in questionnaire design, administration and response.

Table 2: Advantages and limitations of questionnaires in environment health studies.

The use of **structured observations** in environmental health studies has some advantages and limitations as shown in Table 3.

Advantages:
• Information on the physical environment and human behaviour can be recorded
• Observer can 'see' what the untrained eye can miss, as he/she is focussing on the issue.
• Information can be collected on people that cannot take part in interviews, such as babies.
• The information can be checked against other sources, so claims of behaviours in interviews can be checked with observed behaviours
Limitations:
• Observation may not be possible because of social constraints or because the behaviour to observed is rare.
• Behaviour may change due to the presence of the observer.
• Behaviours can be correctly recorded but misinterpreted through the observer.
• It is time consuming and therefore expensive

Table 3: Advantages and limitations of structured observations in environmental health studies.

Table 4 shows the advantages and limitations of using **demonstrations** in health surveys.

Advantages:
• Can be prompted by an interviewer.
Limitations:
• Can be time consuming.
• Result might not be representative of day-to-day practice

Table 4: Advantages and limitations of demonstrations in environmental health studies

In a 'summative' survey, open-ended questions would result in more information than needed and would make it difficult to standardise the outcomes. It would also need more training of the staff doing the interviews. For these reasons open-ended questions will not be considered. As mentioned above, the biggest disadvantage of structured observations is the time needed to make them. For that reason observations will be restricted to spot-observations by the interviewer during the interview. The observations, behaviours or physical characteristics chosen for the survey are only useful if there is a high probability of observing them during the interview. Spot-checks in the case of V21 assessments will observe signs of behaviour rather than the behaviour itself, because it is unlikely that the behaviour will occur during the short period the interviewer is present in the household. An interviewer can always prompt an interviewee for a demonstration if needed.

3.2 Considerations on data collection

For the V21 targets in the survey the BSU is the household and all information will have to be available at the household level. This means that questions about matters such as lining of pits or treatment of tap water are not suitable questions as the kind of information they seek will not always be known at household level.

Several investigations have used observational data but little work has been performed to confirm their validity and repeatability (Boot et al. 1993; Curtis et al. 1993; Ruel et al. 2002). Additional research is needed to assess the validity, reactivity and repeatability of hygiene indicators and composite indices derived from spot-checks in various cultures (Curtis et al. 1993; Kolsky et al. 1995; Ruel et al. 2002).

To make the whole survey widely applicable, the data collected should be as universally relevant and applicable as possible. Also the way the data are collected will determine how widely applicable the method is. Cross-cultural adaptation and testing of questionnaires requires significant time and resources. One way of overcoming this problem is to work with interviewers who specify the interpretation of the questions in the light of local circumstances, and make appropriate observations. This human interface for collecting data might be a good way of increasing the likelihood that the method can be used in most contexts.

It is therefore suggested that the data collection in the survey will be based on:

- interviewer-administered questionnaires in combination with
- interviewer rated spot observations
- demonstrations on interviewers request[1].

The locally selected interviewers will be trained to ensure they obtain valid responses, and will be of the appropriate gender if necessary.

3.3 Consideration on the questions and observations

None of the indicators towards Vision 21 targets can be represented with one single question or observation. To design indicators, different questions and observations will have to lead to a value of the indicator. So not only the indicators have to be determined but also the decision model that leads to the indicator. More questions and observations for one indicator will deal to some degree with non-response (e.g. an observation that was not possible). It also needs to allow for the inclusion of observations such as disposal of stools of small children, which are highly significant in households with small children but completely irrelevant if the household has none.

Questions and spot-observations have the advantage that limited time is needed to collect information. Because the data collected are a surrogate for what we try to measure, it is important that each question and observation clearly states:

- **Why** the question is asked, or the observation is made;
- What will be **concluded** from the answer;
- What were the **assumptions** leading to the conclusions;
- **Remarks** e.g. what to do if the assumptions are proved wrong in a particular survey?

The combination of questions and observations also allows **triangulation** (Almedon et al. 1997; Silverman 2000) to check the validity of outcomes and assumptions. The relationship between the indicator and the different outcomes has to be decided before administering the questionnaire. Examples of a question, an observation and a decision model as submitted in a draft document to the WSSCC can be found in Annex F

After the first peer review of the indicators there was a divide between the major UN organisations (UNICEF and WHO) and other partners of the WSSCC. The issue was whether the questions and observations could be changed to adapt them to local circumstances or were 'set in stone' and should be uniform worldwide.

While the UN organisations did not allow any change to the questions and observations when finalised, smaller organisations wanted a bigger flexibility. The main reason is that smaller organisations like to take surveys as opportunities to collect other data generally of a more formative nature.

The bigger organisations were more concerned about standardisation, comparability and compatibility with data collected in different countries and in the past and future. It was proposed that the survey would comprise three types of data:

[1] At a WSSCC meeting (Geneva June 2002) most organisations present requested to reduce to a maximum the demonstrations in the survey to reduce the contact time between interviewer and interviewee.

1. A **core** set of standardised questions and observations resulting in core data which is minimum data needed to allocate binomial values to the indicator. The relation between the data values and the indicators is determined and there is no room for change unless there is a consensus among WSSCC and its partners.
2. A set of **optional** questions, observations and measurements which allow more rigorous data collection but also will need more resources or skill, such as for testing of water quality. The relation between the optional data and the indicators is also internationally set and change or addition can only be reached through general consensus.
3. **Additional** questions and observations are collected trough locally determined questions observations and other ways of data collection and their use is mainly on a local level.

3.4 Testing the indicators

One of the major problems in measuring hygiene practices is that there is no 'gold standard' (Manun'Ebo et al. 1997) against which validity can be assessed. Moreover questionnaires and spot observations are used to give a value to indicators which are considered practical surrogates for the direct measurement of performance. It is assumed that they express accurately enough the critical aspects of the performance, or in other words that they correlate with reality. It is important that the indicators lead to a result that would also be obtained through more thorough data collection (validity), and that they lead each time to the same result (reliability). In other words, it is important to know how valid the conclusions are when based on the survey data, and also the repeatability of the methodology.

While repeatability or reliability is a primary concern for the indicators, it is not a sufficient condition for validity.

Reliability is here regarded as the degree to which repeated surveys will yield the same results. This could be affected by the validity of the proxy indicators used and is discussed below. It is also dependent on the consistency of the way in which the interviewer or interviewers collect data and classify observations. In the survey used the indicators are kept as simple as possible and the outcome is a set of summative binomial values. This should allow for easier consistency than possible in similar formative surveys. It is still essential to get the right selection and training for interviewers (Kendall et al. 1994).

All the questions and observations used for the indicators are in the private domain and interviewees as well as interviewers might feel uncomfortable with the process. Moreover water, sanitation and particular hygiene practices are private issues and morally loaded. Nobody likes to admit to not washing their hands for example. When selecting ways of validating the survey questionnaire we have to make sure that we use methods that are representing reality. If our validation methods result in similar biases as the survey we might seem to be validating outcomes, as both results are in alignment; but they might not be representative of the real value!

Table 5: Testing of the outcomes for indicators relating to Vision 21 targets 1-3
(Hygiene practice, Access to improved water sources and access to sanitation)

Type of test► ▼Objective	Environ-mental walk	Structured observation (spotcheck).	Pocket charts	Foc. group discussion subsample.	Foc. group discussion population	Key-informant interviews.	Community mapping	Seasonal calendar	Goal of the test	Reliability	Validity	Blind test
Etic 'validation' household level		X	X						Evaluate if the core questions, observations and assumptions are valid	X	X	Y
Etic 'validation' of cluster 'results' of each indicator.	X					X			Evaluate if the results are representative for the targeted population from surveyors perspective.		X	Y
Emic 'validation' of individual interviewing process.				X		X			Appraisal of methods and tools and their relation to trustworthy information.		X	-
Emic 'validation' of cluster 'results' of each indicator.				X			X	X	Evaluate if the results might be representative for the targeted population from their perspective.		X	Y
					X	X			Proclamations of preliminary results from survey and assess population's opinion on them.		X	N
Checking interviewer's bias									Spot observations	X		

3.4.1 Objectives of the validation

The goal of the validation is to see how well these indicators represent reality. Below are some suggestions of how to validate survey questions. The proposed framework of the study is based on **testing the indicators**, but the study design and methodologies should be adapted and finalised by the validation team in the field. Blinded trials mean that the people doing the validation field work are not aware for the outcomes of the households which they visit.

3.4.2 Methods to test the indicators

Table 5 list ways to collect information that can be used for validation of the survey data. The survey is a snapshot of selected information that can be collected in more thorough ways. If for the purpose of validation this information is collected in a subset of the survey sample by a more thorough method, the outcomes obtained could be compared as shown in Figure 2. Only structured observations as data collection will result in data at a household level which is comparable and allow for analyses like shown in Figure 2. For that purpose open ended questions and non-structured observations are not suitable because they give such a wide variety of answers that it is difficult to score them in such a way that they can lead to a value for each indicator.

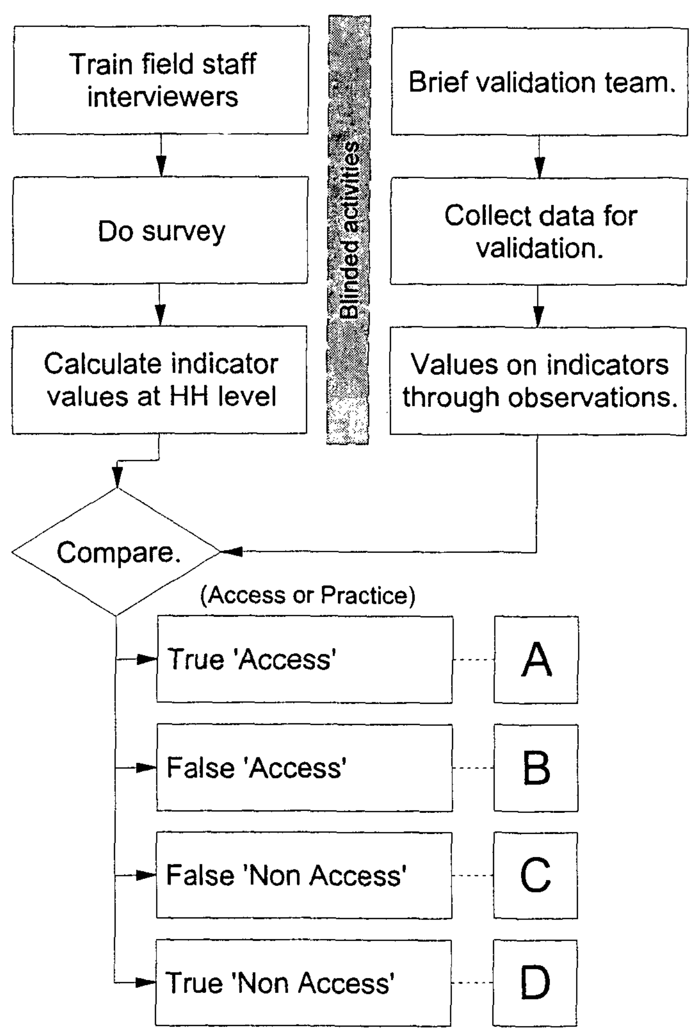


Figure 2: Indicator validation protocol

		'Validation' access or practice status	
		Positive	Negative
Survey status	Positive	*A*	*B*
	Negative	*C*	*D*

Table 6: Overview of validation data

The formulae in Table 7 are used to rate screening programs. They are relatively easy to use; however it is assumed that the validation data is a "gold standard".

$\text{Sensitivity} = \frac{A}{A+C}$	$\text{Positive Predictive Value} = \frac{A}{A+B}$
$\text{Specificity} = \frac{D}{B+D}$	$\text{Negative Predictive Value} = \frac{D}{C+D}$

Table 7: Formulae used for validation

4 Sampling

The goal of the survey is to collect population characteristics in regard to access to water, sanitation and hygiene practices. To obtain information on access and practice it is more economical and practical to select a subset (sample) from the original set of all basic sampling units (the population) of interest. The information we collect (the indicators) are dichotomous. This means that the outcomes follow a binomial distribution. However given a large enough sample size the distribution can be considered normal. The designed sampling methodology to be designed has to be as simple as possible so that sampling, obtaining and analysing results can be done by statistically untrained people.

4.1 Selecting a sampling methodology

Steps in the selection of a sampling process are:

- Define the population and its basic sampling unit.
- Determine the error, relative (ε) or absolute (d) required for the estimates.
- Determine required sample sizes for each sample design taking into consideration the desired confidence intervals.
    - If using cluster sample, determine the design effect (deff), the rate of homogeneity ρ, and adjust the sample size accordingly.
- For each sample design, estimate the field costs
- Choose the design with the lowest field costs.

4.1.1 Basic sampling unit

Generally the population targeted by the survey will be determined by existing administrative boundaries and geographical locations. "Population" in this document is used in its statistical sense as the collection of all the basic sampling units (BSUs). The basic sampling unit is the individual member of the population whose characteristics we want to measure. For Vision 21 targets 1-3 (see Annex A) the basic sampling unit is the household. Research findings suggest that as neighbourhood levels of faecal contamination improve, the conditions and practices within households become more important. This means a move away from the traditional, engineering approach to public health (Cairncross et al. 1996). It brings the focus towards *private* health at household level. The concept of the *domestic domain* encompasses the decisions and actions taken at household level and their relation to environmental health, and is distinguished from the public domain in which the intervention of public authority is required to prevent disease transmission. This model acknowledges the importance of household practices and behaviour. A household-centred approach to environmental health has also been advocated by. (DFID 1998)
Households are universal and relatively easy to identify. This makes them suitable to be used as basic sampling unit (BSU). However choosing the household as a basic sample unit makes the outcome a percentage in terms of households. All distinction between various types of household, such as man alone or woman and children, will be lost for analysis. This means also that gender issues will be lost.
The term "household" may be interpreted according to local conditions; however a convenient definition could be "those whose food is prepared by the same person" (Bennett et al. 1991) or "those who slept in the same building last night". This definition might still pose a problem for the increasing amount of single (mostly man-only) households in urban slums or rough sleepers without a physical 'address'.
It might be necessary to verify if the data obtained at household level is representative of the population before inference is made from the obtained results.
The problem in sampling households is that aspects such as sanitation in public places are left out of the equation. Including it in the same survey would be difficult as the evaluation is different and a weighting factor to include it in the same statistics would be difficult to calibrate.

4.1.2 Representation of the basic sampling unit

Who in the household will give the information that is most representative for the household? Women have been traditionally at the 'practical' day-to-day centre of the household. They are usually involved in the collection of water, preparing the food, taking care of the children and maintaining cleanliness in and around the dwelling. So they seem to be the most suitable candidates to interview. This assumes that there is in most cases a 'normal' family constitution. In some cultures, interviewing women might not be straightforward. With so many responsibilities, women might not always be available to give information; this might increase the non-response rate. It is suggested that the person involved in the cooking, cleaning and collecting of water for the household is the person to interview. It is assumed that this person will generally be the 'woman of the house'.

4.2 Sample size determination for testing the indicators

The simple random sample is the gold standard in sampling, and if a simple random sampling would be used to determine the proportion in the target population the sample size needed can be calculated with the following equations.

	Exact	Approximate
Sample size for proportions with relative deviation	$n \geq \frac{Z^2 N P_x(1-P_x)}{(N-1)\varepsilon^2 P_x^2 + z^2 P_x(1-P_x)}$	$n \geq \frac{z^2(1-P_x)}{\varepsilon^2 P_x}$
Sample size for proportions with absolute deviation		$n \geq \frac{z^2 P_x(1-P_x)}{d^2}$

(Lemeshow et al. 1990; Levy et al. 1999)

- z Reliability coefficient or amount SE away from the mean
- n Sample size
- N Population
- ε Relative deviation (%) of the result, εP_x= d, Absolute deviation of the result
- d Absolute deviation (% points) of de result or PRECISION
- P_x Unknown population proportion for sampled variable

Equation 1: Sample sizes in simple random samples

The maximum sample size needed for a simple random sample with a precision of 10 % points would be when $P_x = 50\%$. The sample size needed would be 96 basic sampling units as calculated in Equation 2.

$$n = \frac{z^2 p(1-p)}{d^2} = \frac{1.96^2 \times 0.5(1-0.5)}{0.1^2} = 96$$

- z=1.96 Reliability coefficient or amount SE away from the mean
- d=0.1 Absolute deviation (% points) of de result or PRECISION
- n Sample size

Equation 2: Simple random sample size calculation

Simple random sampling is unpractical and costly for use in a large scale population survey because individual sampled units can be many kilometres apart. That is why multistage cluster sampling was developed. There are many different designs for multistage cluster sampling surveys.

4.3 Multistage cluster sampling.

Multistage cluster sampling splits the selection process into multiple stages; at each stage different selection procedures can be used to select various subsets until the BSUs are selected. A commonly used and simple design for a two stage cross sectional survey is to select the primary sampling units (PSUs) with a 'probability proportional to size'. The second stage is typically a random (or pseudo random) sampling of equal size in each of the selected PSUs. It can be proven mathematically that in this way all the BSUs have an equal non-zero chance of being selected. This type of sampling is referred to as a self weighted sample or an equal probability of selection method (EPSEM) because no weights have to be given to the individual samples or PSUs. The number of samples selected in the cluster, also called take, are for each cluster the same which results in a simple self weighted sampling method.
There are however limitations to this way of sampling:

- The measure of size used for the PPS sampling of PSUs can be different from the BSUs. E.g. if the BSU is children between 12-36 months the number of children for each PSU might not be known. The number of households could be taken as a measure of size on the assumption that there is a linear proportional relation between the number of BSUs and the measure of size.
- Often, estimates of size are used rather than the known size, which can reduce the statistical validity of the sample result.
- If the survey population is not homogeneous (it rarely is) there is the possibility that some PSUs group BSUs with a particular value. This clustering increases the possibility that the survey values are not representative for the population. This uncertainty, expressed as the design effect, will increase the variance of the survey measurement.

The increase in the variance is expressed by a factor D called the design effect (deff) ratio.
Including the design effect the equation for the standard error becomes:

$$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}\hat{D}} \quad \Rightarrow \quad n \geq \frac{p(1-p)}{s^2}\hat{D}$$

(Bennett et al. 1991)

In a cluster design the sample size has to be increased with the design effect to result in a similar standard error as a SRS

s	standard error SE	D	design effect (deff)	p	proportion
$\hat{s}$	estimated SE	$\hat{D}$	estimated deff.	n	sample size

Equation 3: Standard error and sample size in cluster sampling

4.3.1 Design effects and rate of homogeneity

Among diverse selection procedures the chosen selection method has by far the greatest effect on both the variance and the cost (Kish 1965). Samples obtained through cluster sampling cannot be directly calculated with Equation 1. Cluster samples are characterised by a higher homogeneity than simple random sample in the same population which tends to increase the variance of the sample. Generally cluster samples have a larger standard error which requires them to have a larger sample to obtain similar standard errors as if simple random sampling were used. The factor expressing this difference is called the design effect (Deff). *Design effect* are defined as '*the ratio of the actual sampling variance to the variance of a simple random sample with the same number of units*' (Yates 1981).

$$Deff = \frac{\text{actual sample variance}}{\text{variance of a random sample with the same number of basic sampling units}} = \frac{V_{cs}}{V_{srs}}$$

Deff. design effect

V_{cs} Variance of the cluster sample

V_{srs} Variance of a simple random sample with the same number of BSUs in the same population.

Equation 4: True design effect

In reality we can only work with information from our sample. Moreover we do cluster sampling to avoid having to take simple random samples from the population. Without a simple random sample for comparison, the true design effect would be only a theoretical concept. However, in practice, the design effect is calculated as '*the ratio of the actual sample variance as cluster sample* to *the variance of the* *same sample calculated as if it were a random sample'*. This gives us an idea of the possible true design effect, which in real sampling will stay unknown.

$$deff \cong \frac{\text{cluster sample variance}}{\text{variance of the same BSUs calculated as a simple random sample}} = \frac{\hat{V}_{CS}}{\hat{V}_{srs}}$$

Deff: design effect

V_{cs} Variance of the cluster sample

V_{srs} Variance of the same sampled BSUs calculated a simple random sample.

Equation 5: Practical calculation of design effect

Design effects express partially the clustering of the measured characteristic in the population. This effect can be reinforced or attenuated by the clustering of the samples. Design effects are empirically obtained through a survey of a similar design. What exactly has to be similar is not clear from literature but the take size and the number of clusters seem important factors in the design of a two stage cluster sampling. To have a better insight into the 'typical' design effects for access to water and sanitation, data from some existing DHS and MICS data sets were analysed (see selection 4.3.2.) The introduction of a generalised intra-cluster correlation coefficient , termed "rate of homogeneity" (*roh*) by Kish (1965) aims at removing the effects of average cluster sizes in the comparison of results across different variables and population domains.

$$\hat{deff} = \frac{V_{CS}}{\hat{V}_{srs}} = \frac{V^2(\hat{\bar{X}})}{\hat{S}^2_{srs}/m} = 1 + \rho(\bar{b} - 1)$$

(du V Florey 1993; Kish 1965)

$$\hat{\rho} = \frac{\hat{Deff} - 1}{\bar{b} - 1}$$

(Kish 1965)

deff: design effect

V_{cs} Variance of the cluster sample

V_{srs} Variance of the same sampled BSUs calculated a simple random sample.

S_{srs} Standard error of the same sample, calculated simple random sample.

$\hat{\bar{X}}$ Estimated mean of the measured variable

S_{cs} Standard error of the cluster sample

m sample size

ρ rate of homogeneity *roh*[1]

b take or cluster sample size

Equation 6: Design effects and rate of homogeneity *roh*

Montanari (1993) looked at how valid ρ was in expressing the rate of homogeneity within the population rather than within the sample. He found that the ρ will only do this under certain conditions and states that there might be better indicators for this. To do this he introduced a variable *k* as shown below.

$$\hat{deff} = (1 - k)\left[1 + \rho(\bar{b} - 1)\right] \quad \text{with} \quad k = A_{D0} + A_{D1} + A_{D2} + B_{D1} + B_{D2}$$

(Montanari 1993)

Equation 7: Variable 'k' in the relation between deff and ρ

The formulae for each of the factors determining *k* are relatively complex and theoretical but Montanari (1993) states that under particular circumstances k becomes negligible and Kish formula becomes valid.

[1] In literature the abbreviation 'roh' for rate of homogeneity and 'rho' for de Greek letter 'ρ' are both used to denote the generalised 'intra-cluster correlation coefficient' ICC.

For dichotomous variables these conditions are:

- *Allocation of PSUs to the strata is proportional to the stratum sizes.*
  The designed survey methodology will only apply explicit stratification which means that for each stratum (e.g. Rural, Urban) a separate survey has to be done. This means that each survey only has one strata and this condition is in our methodology fulfilled at all times.
- *The PSU selection probabilities are proportional to the domain sizes and the population is the domain.* In the survey no data is provided for the analyses of different domains so the population is the domain, which together with PPS selection of the PSUs provides a partial condition for k to become negligible.
- The last condition to use Kish's formula is that the sampling in the survey is self-weighted (EPSEM) which it is.

In our survey design conditions for $k \approx 0$ will therefore be fulfilled. This means that we will use Equation 6 as initially formulated by Kish (1965). Moreover including k would increase the statistical complexity dramatically without clear benefits. To keep the statistics in the survey method simple we will follow the suggestion of Kish (1965) to use empirical studies to improve the estimates of *deff* using the rate of homogeneity ρ as defined in Equation 6.

Thomsen et al. (1986) looked at the effect of using an out-of-date measure of size in determining ρ and noticed that the differences can be considerable. This is not so important for our survey methodology and its analysis but very important for the abstraction of ρ from new or existing surveys where the measurement of size used might not be correct. The ρ will vary between a maximum of one for high design effects and 0 for none. Roh can occasionally become negative if clusters result in a lower variance than in simple random sampling. This situation is however exceptional.

4.3.2 Existing data collection, MICS, DHS

To estimate design effects and the rate of homogeneity ρ some DHS and MICS datasets have been analysed. The datasets used were DHS for the Dominican Republic 1994, DHS Kenya 1998 and MICS Moldavia 2000. After the cleaning of the data a new binary value was created representing access to an 'improved water source' based on the information available in the DHS and MICS data sets,. As water, sanitation and hygiene data are collected as an extra add-on to these surveys the non-response to these questions is high. This non-response is considerably higher for sanitation and hygiene behaviour. The DHS for the Dominican Republic 1994 was chosen to calculate ρ as out of all the datasets analysed it had the lowest non-response ratio (13%, n=8830) for access to water. The data set was the only one that contained information on drinking and non drinking water, which made possible to create better access indicators. Water is expected to have the highest design effects because it has a geographical determinant (the distance to the source) to it. This is demonstrated in Figure 3 by an example of an EPI sampling (explained in section 4.6.1 on page 24). This geographical determinant is the reason why high design effects are expected for the access to water indicator. The dataset yielded a design effect ratio for access to water indicator of 7.5. This result in a rate of homogeneity:

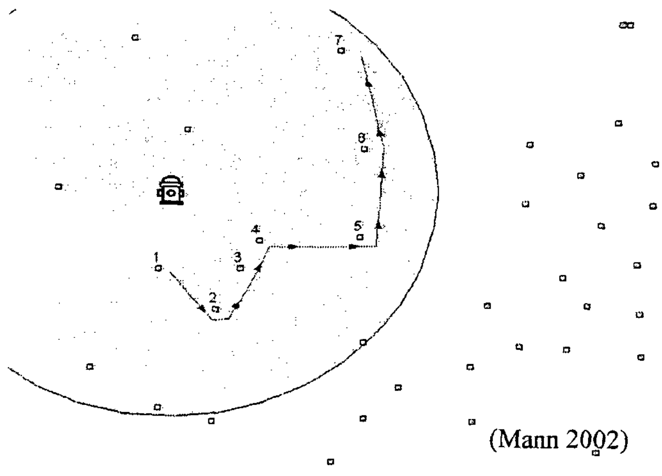


Figure 3: Geographical determinant to 'access to water'.

$$\hat{\rho} = \frac{\hat{Deff} - 1}{\bar{b} - 1} = \frac{7.5 - 1}{22 - 1} = 0.3095 \cong 0.31$$

Deff. design effect

ρ rate of homogeneity *roh*

b take or cluster sample size

To determine how typical this is, more data sets will be analysed in the future of the project. To have a better idea of the size of design effect and rate of homogeneity and how it relates with other cross sectional survey data, some other studies where consulted of which one (Qaba 1999) is added in Annex B for comparison.

The DHS sampling methodology.

One of the criteria for the rate of homogeneity to be representative for population in the survey design is that it uses similar definitions for the targeted population.

The general policies for the DHS surveys are (Macro 1996):

- 100% national coverage (some conditional minor exclusions allowed)
- Probability sample. That is a sample in which each BSU has a known and non-zero probability of being selected.
- Self weighted sample, equal probability of selection method (EPSEM)
  Sample in which each BSU has an equal probability of being selected
- Use of pre-existing sample frames (if available and adequate)
- Simplicity of design

DHS is designed for a target sample size of 5000 to 6000 women age 15-49 while a max of 10% under sampling is considered due to under coverage at the household mapping and listing stage or non response. DHS allows implicit stratification by allowing the distinction of five to six regions with around 1000 sampled in each region.

The sample size in each cluster in the DHS surveys is called 'take' and for general purpose DHS the take is '20' for urban and 30-40 for rural women.

Non sample frame sampling (without household listings) is not accepted in DHS as it is considered a false economy.

Because the Demographic Health Survey (DHS) concentrates on demographic and reproductive data it targets women aged 15-49. It is difficult to evaluate how the concentration on this group might influence the rate of homogeneity for access to water in the new survey method.

4.4 Sampling with a sample frame

For the testing of the indicators, only sampling with a sample frame will be used. For this only places in which a recent DHS or MICS surveys was held were considered so that the same sample frame can be used. But even for these situations the number of clusters and required sample size has to be calculated. Equation 8 allows the calculation of the number of cluster needed when p, d, b and deff are known.

$$c = \frac{p(1-p)deff}{s^2 b}$$

(Bennett et al. 1994)

c= cluster size

p= expected proportion

s= standard error

d= absolute error in % points

z = reliability coefficient

b= average take

$d = 0.1$ or 10% points

$z = 1.96$ or 95% confidence level

$$s \approx \frac{d}{z} = \frac{0.1}{1.96} = 0.051$$

if $p = 0.5$ or 50% $\Rightarrow$

$$c = \frac{p(1-p)}{s^2 b}\left(1 + \rho(b-1)\right)$$

$$c = \frac{0.5(1-0.5)}{0.051^2 b}\left(1 + \rho(\bar{b}-1)\right)$$

$$c \cong \frac{96(1-\rho) + 96 \cdot \rho \cdot \bar{b})}{\bar{b}}$$

$$c \cong \frac{66 + 30\bar{b}}{\bar{b}}$$

Equation 8: Cluster size

We can calculate s for 10% points confidents interval at the 95% confidence level (z=1.96) as shown. Using the worst case value of 50% as expected proportion and replacing the deff in Equation 8 by Equation 6 which with the other values result in the shown equation.
There is no optimal combination to be found on the basis of this equation as shown in Figure 4.

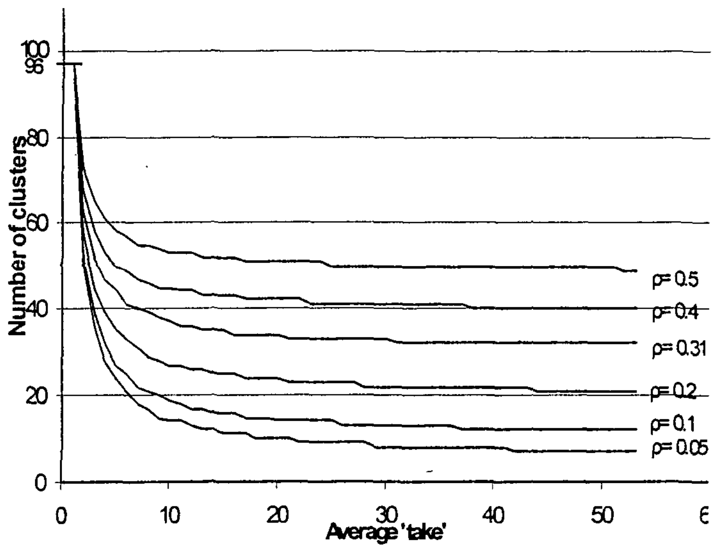


Figure 4: Number of clusters versus average 'take' size

To calculate the max and minimum number of clusters, the limits for this function are calculated for ρ = 3.1

$$\lim_{\bar{b}\to\infty} \frac{66 + 30\bar{b}}{\bar{b}} = 30$$

$$\lim_{\bar{b}\to 1} \frac{66 + 30\bar{b}}{\bar{b}} = 96$$

For any value of ρ the equation becomes as shown in Equation 9

Minimum number of clusters	$\lim_{\bar{b}\to\infty} \frac{96(1-\rho) + 96 \cdot \rho \cdot \bar{b}}{\bar{b}} = 96\rho$
Maximum number of clusters	$\lim_{\bar{b}\to 1} \frac{96(1-\rho) + 96 \cdot \rho \cdot \bar{b}}{\bar{b}} = 96$

Equation 9: Min. and max number of clusters

The maximum number of clusters is occurs when the take size is 1 which corresponds with a simple random sample while the minimum is dependent on ρ. There is however no optimal relation between number of cluster and the take size.

4.5 Cost factor in the cluster-take relation

One of the main factors for cluster sampling is the cost factor. Introducing this cost factor into the equation as shown in Figure 5 shows that an optimum can be reached. To only include one extra variable on cost in the equation cost ratio between the 'extra' cluster cost and the 'cost of sampling one BSU' was used. The cluster cost is the extra cost a cluster adds to the survey budget on top of the cost of sampling all the BSUs in that cluster. The (cluster cost)/(cost of sampling one BSU) ratio including an example is further explained in Annex D. The graph in Figure 5 is not fluent because only integers where used in the formulas. The graph also shows clearly that the optimal sample size can change significantly with slight variations of ρ. To get a better understanding of the cost and its relation with the sample size the cost ratio between cluster and sample size are plotted in Figure 6 For a given ρ=0.31.

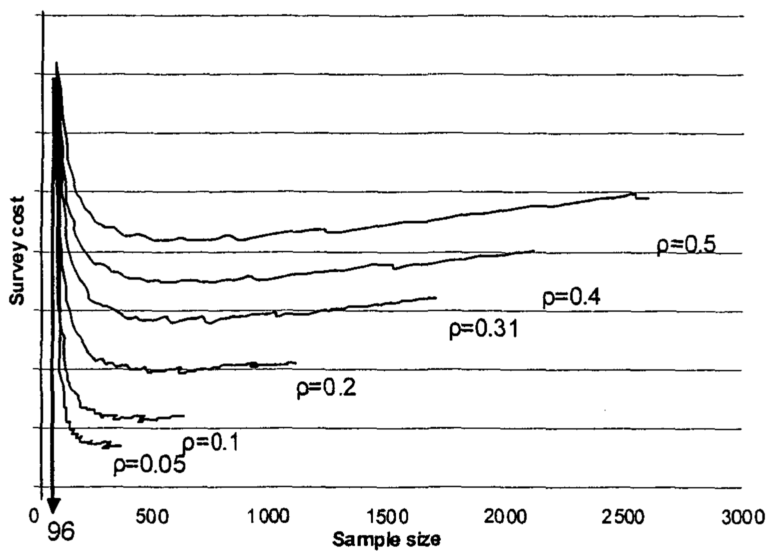


Figure 5: Optimal cost versus sample size

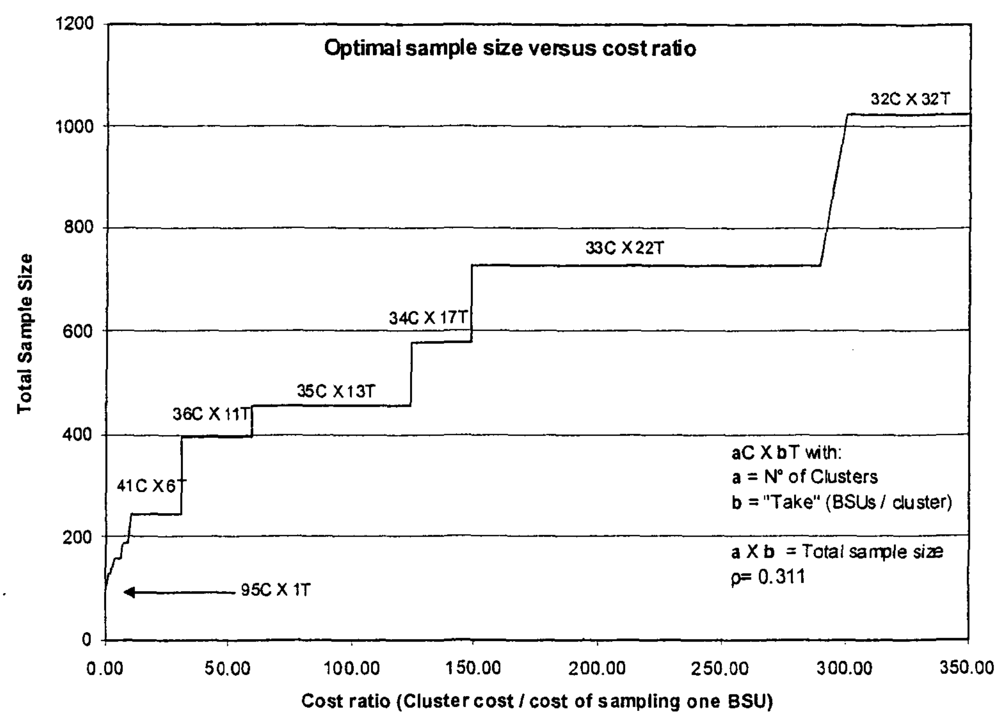


Figure 6: Optimum sample size versus cost ratio

The formulas used are based on sampling 'with replacement'. They can only approximate 'sampling without replacement' when the cluster population sizes are much larger that the take. In practices this condition is fulfilled when the cluster population is at least 10 times larger than the take.

Where the lines in the graph are vertically there are two optimal combinations. At a cost ratio of 148 the combinations 34C X 17T and 33C X 22T will give the same overall cost for two distinctive different sample sizes. Cost In the calculations is normalised at the cost of surveying a single sample. It shows that despite the cost expressed in cost per unit sampled, the two different sample sizes are significantly different.

Survey cost (in samples) = sample size + cost of cluster (in samples)

$$\text{sample size} = 34C \times 17T = 578 \quad \Rightarrow \quad 578 + 34C(148) = 5610 \text{ (cost in nr. of samples)}$$
$$\text{sample size} = 33C \times 22T = 726 \quad \Rightarrow \quad 726 + 33C(148) = 5610 \text{ (cost in nr. of samples)}$$

Any cluster/sample cost ratio over 350 results in the same outcome of 32 clusters of 32 BSUs each.

4.6 Sampling without a sample frame

In low income countries information for making sample frames is not always readily available or kept up to date. The cost of creating a sample frame can be a heavy burden on a data collection budget. For these reasons some alternative cross sectional methodologies were designed to obtain samples without a sample frame.

4.6.1 EPI sampling method

The best known technique is the EPI method developed for and named after WHO/Unicef's Expanded Program for Immunisation (EPI). It is based on a sampling technique originally developed in the United States (Serfling et al. 1965) and later adapted for the smallpox eradication campaign in West Africa. (Henderson et al. 1973; Henderson et al. 1982; Levy et al. 1999). The method selects 30 clusters from the PSU by probability proportionate to an estimated size (PPES) of measure like households or populations. In each of the clusters 7 children are selected. This is done by going to the estimated centre of the cluster, spinning an object to choose a random direction, and listing all the households in that given direction. From that list, one household is selected randomly which is the first household that will be checked if it fulfils the criterion to be included into the sample. From thereon the house with its entrance closest to the last household is checked until 7 children between 12-36 months have been selected in the primary sampling unit. The spread is obtained because in surveys such as for polio vaccination coverage, only children between 12-36 months are being considered which are not present in all households.

Various criticisms of the EPI method and suggestions for improvements have been made (Bennett 1993; Brogan et al. 1994; Turner et al. 1996). The major critic is that the EPI method has a tendency of clustering around the first house selected for sampling. This 'second stage clustering' seems to be compensated for by using correct values for the design effects when using the EPI method.

Sampling methods for cross-sectional sampling methods without a sampling frame like EPI move away from random sampling. These methods have as goal to simulate a simple random sample as well as possible so that statistical inference is still possible. The components and protocols which make up the sampling method (like choosing the nearest house) are difficult to integrate in simple mathematical formulae and their effect on potential bias is difficult to treat analytically. For that reason these sampling methods are best tested in a real world situation or in a computer simulation.

4.6.2 Other sampling methods

For the moment only two sampling methods are taken into consideration in this project. The first one is using a selection criterion like children between 12-36 months in the EPI method, which would increase

the geographical spread of the sample. As such criteria might also result in a biased sample an adapted EPI method by selecting every n^{th} household was preferred as tested by (Bennett et al. 1994)
The second method is a random geographical method in which a random geographical point is chosen in the PSU and the closest household to it is selected. The next household selected is the closest to another random geographical point.
While some preliminary result indicates that EPI n^{th} seem still to have a bias for higher population densities the random geographical point method is expected to have exactly the opposite.
The second method will need the use of GPSs which will increase the cost of surveying. This seems to be a contradiction for a method that aims to decrease survey cost. Both financial advantages and disadvantages will have to be watched carefully to assess how far the latter method is useful.

4.6.3 Computer simulations

For practical and financial reasons, computer simulations are chosen to test these sampling methods. This approach has been used in the past to test the EPI method (Bennett et al. 1994; Lemeshow et al. 1985) and some alternative methods (Bennett et al. 1994). While both approaches to simulating the EPI and alternative methods had advantages and disadvantages, they proved essential for the development and acceptance of sampling methods without a sample frame.
Building computer simulations has a major problem that the spread and clustering of the property to be measured often has to be simulated due to lack of real data such as georeferenced census data. The model of spread has often a big influence on the sampling. With water there is in most cases an important geographical component (distance from water sources) which makes this modelling more likely to be realistic.

To do computer simulation a simulator had to be designed. The idea was to design one in an existing GIS packages as they are designed to examine geographical data. This would allow using existing data in the simulator. Esri's commercial ArcGIS™ products were chosen as a development environment and the development language in ArcGIS™ is Visual Basic™ for applications (VBA) and compiled Visual Basic routines.
This choice was based on the products available on the LSHTM's network and the centres willing to cooperate in the development. The collaborating centre of choice for the development was UCL centre of geomatic engineering because of its proximity to the school. One student, Gareth Mann, successfully made his MSc dissertation on the development of the first version of the simulator. Two years research funding will be sought for further development of the simulator. is a three dimensional view of a cluster in which 665 iterations of 7 samples where taken. The size of the 'buildings' indicate the number of times a sample was selected. The tall buildings in the middle of the EPI sample indicate over 50 times sampled. EPI had a large number of households not sampled.

the geographical spread of the sample. As such criteria might also result in a biased sample an adapted EPI method by selecting every n^{th} household was preferred as tested by (Bennett et al. 1994)
The second method is a random geographical method in which a random geographical point is chosen in the PSU and the closest household to it is selected. The next household selected is the closest to another random geographical point.
While some preliminary result indicates that EPI n^{th} seem still to have a bias for higher population densities the random geographical point method is expected to have exactly the opposite.
The second method will need the use of GPSs which will increase the cost of surveying. This seems to be a contradiction for a method that aims to decrease survey cost. Both financial advantages and disadvantages will have to be watched carefully to assess how far the latter method is useful.

4.6.3 Computer simulations

For practical and financial reasons, computer simulations are chosen to test these sampling methods. This approach has been used in the past to test the EPI method (Bennett et al. 1994; Lemeshow et al. 1985) and some alternative methods (Bennett et al. 1994). While both approaches to simulating the EPI and alternative methods had advantages and disadvantages, they proved essential for the development and acceptance of sampling methods without a sample frame.
Building computer simulations has a major problem that the spread and clustering of the property to be measured often has to be simulated due to lack of real data such as georeferenced census data. The model of spread has often a big influence on the sampling. With water there is in most cases an important geographical component (distance from water sources) which makes this modelling more likely to be realistic.

To do computer simulation a simulator had to be designed. The idea was to design one in an existing GIS packages as they are designed to examine geographical data. This would allow using existing data in the simulator. Esri's commercial ArcGIS™ products were chosen as a development environment and the development language in ArcGIS™ is Visual Basic™ for applications (VBA) and compiled Visual Basic routines.
This choice was based on the products available on the LSHTM's network and the centres willing to cooperate in the development. The collaborating centre of choice for the development was UCL centre of geomatic engineering because of its proximity to the school. One student, Gareth Mann, successfully made his MSc dissertation on the development of the first version of the simulator. Two years research funding will be sought for further development of the simulator. is a three dimensional view of a cluster in which 665 iterations of 7 samples where taken. The size of the 'buildings' indicate the number of times a sample was selected. The tall buildings in the middle of the EPI sample indicate over 50 times sampled. EPI had a large number of households not sampled.

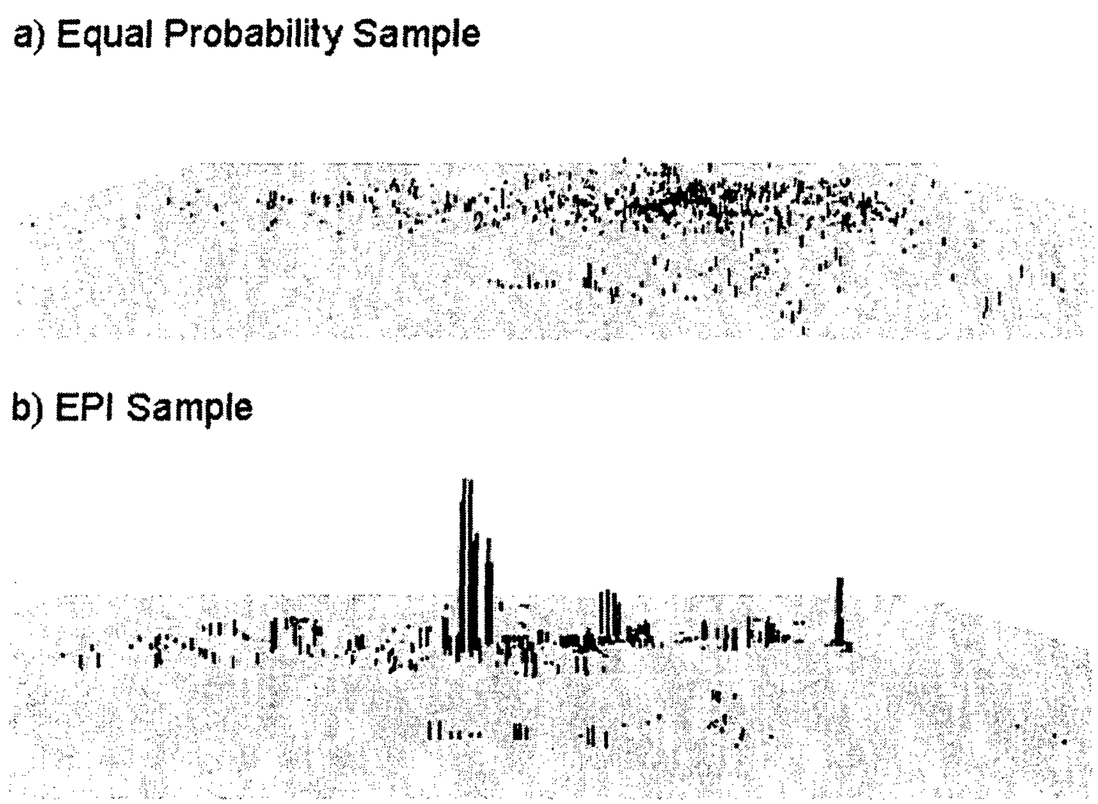


Figure 7: 3D view of the sampling result after 665 times sampling of the same PSU.

The above figure shows clearly that the EPI has a strong clustering effect. This is not necessary a problem if it does not lead to biased results. However with access to water this geographic clustering might be a problem as demonstrated in Figure 3 on page 21. Apart from the cost and the analytical capacity of computers there is another advantage of computer simulations. The data does not have to be compared with simple random sampling but can be compared with the 'real' population figures in the simulator.
It is hoped that a complete georeferenced population data set can be obtained, such as local census data, for testing purposes.

5 Results of preliminary work

This is a brief recapitulation of the work already done as an introduction to the next paragraph going in detail trough the work ahead.

5.1 Indicators

a. At the moment a survey questionnaire including spot observations has been designed. It was reviewed in a WSSCC meeting in June 2002. A reviewed draft as attached in Annex A went out for peer review on the 15th of August. Feedback on this document will be compiled at the beginning of November to a final draft.

b. A first pre-testing of the questionnaire was done June-August 2002 in Kosovo. The few results available for the moment will be incorporated in the next versions of the questionnaire.

c. A scoring sheet for structured observations is being finalised but a scoring protocol to compare it with the survey result still needs developing.

5.2 Sampling

a. Some brief analyses of existing DHS and MICS surveys was done to obtain a value for the rate of homogeneity for water, sanitation and health behaviour.

b. Based on the *roh* obtained the number of clusters and the sample size needed to test the indicators was calculated.

c. A computer simulator was designed by an MSc student (Gareth Mann) at UCL's department for geomatic engineering (see also Annex G), but was unfortunately not entirely designed as requested. The program works but some parts still have to be checked for their statistical validity. With some small changes it should be possible to do some testing at the cluster level. The simulator is at the time running too slowly to run many iterations to obtain full surveys (e.g. 1000 iterations) as all simulations are run at cluster level and then merged. See also Annex H for details on the simulation process.

6 Future plans

Dates mentioned in brackets are the planned end dates or periods for the work described. A full time scale of further work can be found in Figure 8 on page 30.

6.1 Indicators

6.1.1 Finalising the indicators

a. Incorporate all the remarks from the last peer review into a new reviewed document (8/11/2002).

b. Finalise the questionnaire and the decision models. (8/11/2002)

6.1.2 Testing indicators

a. Finalise the structured observations marking sheets.

b. Determine a scoring mechanism for structured observations leading to indicator values.

c. Selection of best statistical tool to evaluate indicators.

d. Testing of the indicators will be done from now until the end of January by:

   i. WaterAid and partners in Bangladesh

   ii. UNICEF in Myanmar

   iii. Department of Water and Affairs and Forestry in South Africa

   The following three organisations have expressed an interest but no recent news has been received on the status of their cooperations.

   iv. Earth forever in Bulgaria

   v. NetWas in Kenya

   vi. CINARA in Nicaragua

The field test of the survey will be done in 32 clusters of each 32 households. The survey will be done by a group of people trained on the spot for doing the survey. After this a subset of random selected households included in the survey will be resampled. These will be visited by trained observer who will do a more thorough evaluation of these household to be compared to the survey outcomes as shown in Figure 2 on page 14. These will be combined with other types of evaluations as shown in Table 5 on page 13. These include evaluation such as emic evaluation of the methodology trough focus groups and key informants discussion groups. This would give an inside in how the interviewee experienced the evaluation and how that affected their responses.

a. Preliminary analyses of survey data for presentation at the World Water conference in Kyoto March 2003. (Dec.2002, Jan & Feb 2003 end 28/02/2002)

   This analysis is to present preliminary survey results but is not concerned with the validation of the indicators. The latter will be done in March April and May 2003.

b. Analysis of indicators in the different surveys to analyse the methodology. (March April and May 2003)

   Methods suggested for this are:

   i. Logistic regression of the question and observations in relation to the indicator outcomes to analyse the importance (weight) of each of them in relation to the outcome. This will also determine where the most effort will be needed to get that particular question or observation right or which can be omitted as it does not contribute at all toward the indicator.

   ii. Examine the agreement between multiple methods to measure the same outcome based on statistic formulae as shown in Table 7. See what has to be changed to improve the survey and increase the agreement between both methodologies.

6.1.3 Analysis of existing data sets

Around six existing data sets DHS and MICS will be examined between May and October 2003 mainly to study the rate of homogeneity and the design effects (May-Oct. 2003)

6.2 Sampling methods without sample frame

6.2.1 Designing the sampling method

As mentioned before only two methods are being considered:

i. the EPI n^{th} method (in which only each the n^{th} household is included into the sample) and

ii. the random geographical point method

The EPI n^{th} method, and some variations, has already been programmed into the sampling simulator while the random geographical point method not yet. The limitation in testing different methods and variations is mainly due to the speed of the simulator which is far too slow to consider more than two options.

6.2.2 Testing the sampling method

a. Improve and further develop the computer program for testing various sampling schemes. This will be done at UCL on a 2 year research grant to be introduced end of November 2002.

b. ANOVA and regression analyses based on cluster and sample data collected in the sampling data set generated by the simulator. The program allows data to be collect for each sample and for each cluster, similar to those published by.(Lemeshow et al. 1985)

   Information considered for this analysis are:

   i. Household density at household level
      (The number of households that are in a set radius around the household.)

   ii. Household density spread (z score) of the household densities at cluster level

   iii. Number of sources at cluster level

   iv. Centralised or decentralised sources (water access pattern) at cluster level.

   v. Access coverage.

   More information on the different steps involved in the sampling simulation is shown in Annex H.

a. If different sampling methods deliver best results under different conditions, find an easy way to analyse these conditions in a real world situation.

b. Field test of the whole methodology.

   To do this a location has to be found in which recent census was held in with detailed data was collected on water and sanitation. The results of the census data would allow comparison with the result obtained in the survey.

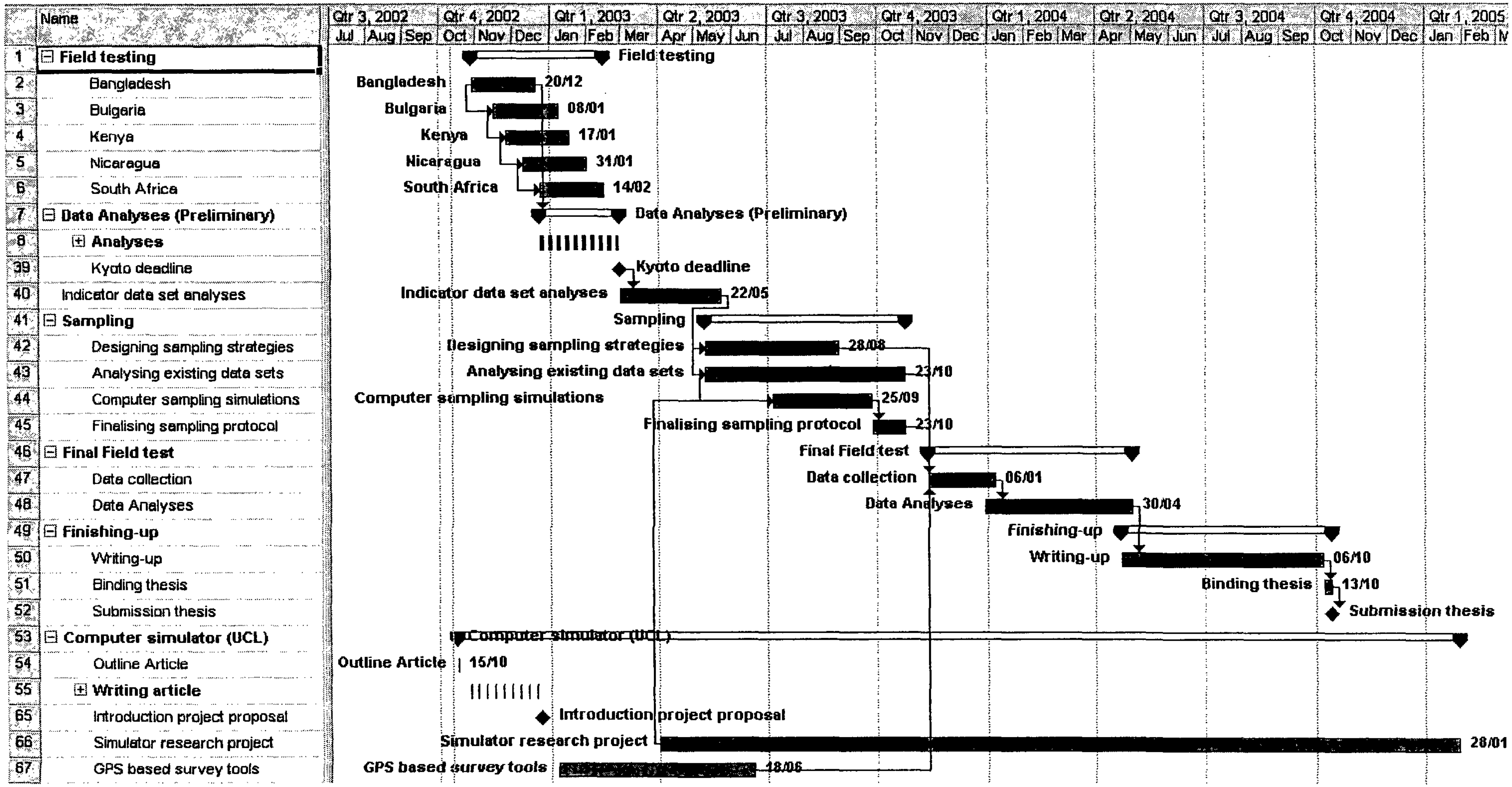


Figure 8: Planning and time line of further activities

Annex A Vision 21 Targets

	Vision suggested targets for 2015	for 2025
1	Universal public awareness	Good hygiene practices universally applied
2	Percentage of people who lack adequate sanitation halved	Adequate sanitation for everyone
3	Percentages of people who lack save water halved	Safe water for everyone
4	80% of the children educated about hygiene	All primary school children educated about hygiene
5	All schools equipped with facilities for sanitation and hand-washing	
6	Diarrhoeal disease incidence reduced by 50% (indicator considered for health sector)	Diarrhoeal disease incidence reduced by 80%

Table 8: Vision 21 targets for 2015 and 2025

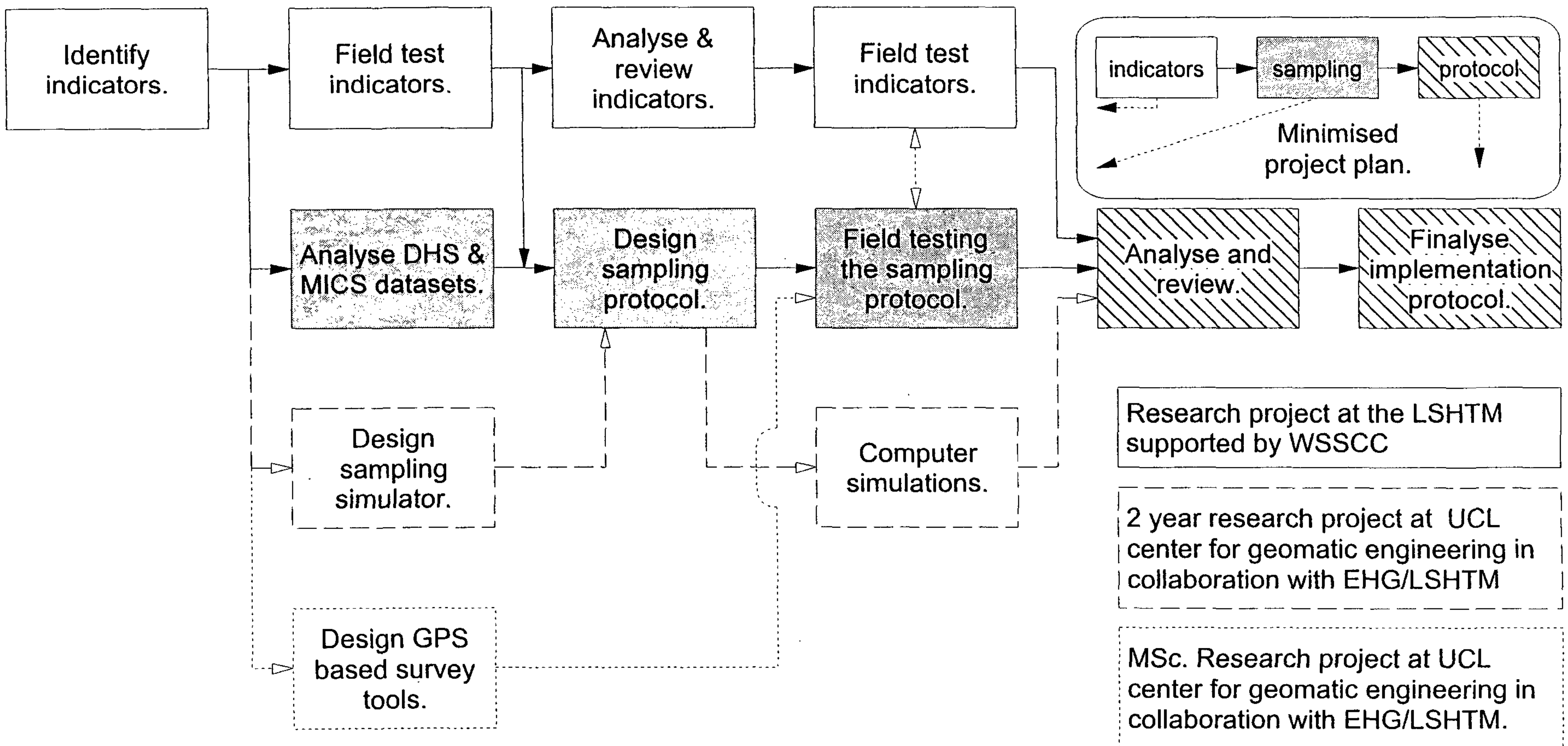
Annex B Examples of design effects of other variables.

The survey data shown below was collected through a two stage cluster sample design of which the PSUs where selected PPS based on census data. The sampling in the cluster was a systematical 'random' sample from the census listing. A take of 25 sample units was taken from each of the 360 selected PSUs.

Variable and description	n	Average deff	Average roh
ACCOMODATION			
Type of house	24427	15.04	0.5907
Roof material	24387	14.59	0.5678
DEMOGRAPHIC			
Gender	24452	2.26	0.0188
Race	24452	16.21	0.6399
SOCIO-ECONOMIC			
Employed	13647	5.36	0.2206
HEALTH			
Consultation	2475	1.31	0.0046
AGRICULTURE			
Type of land ownership	2829	10.71	0.1443

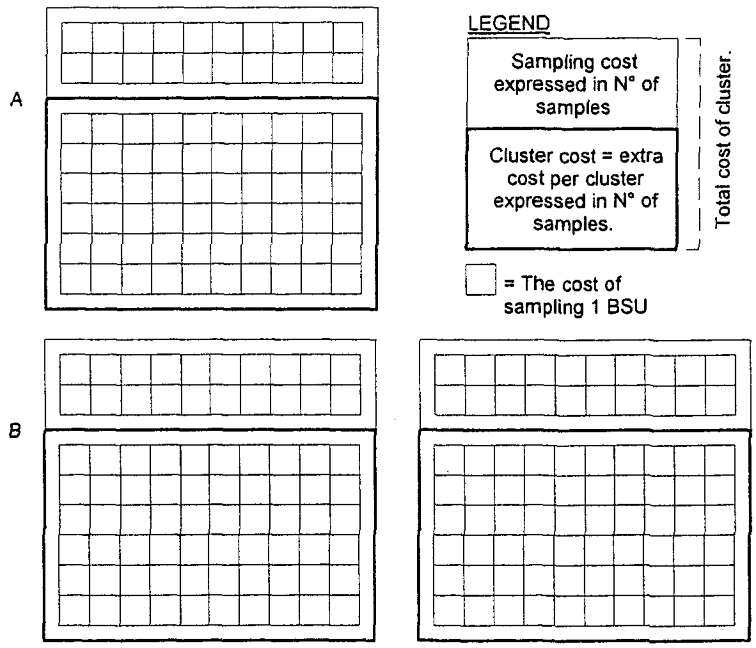
(Qaba 1999)

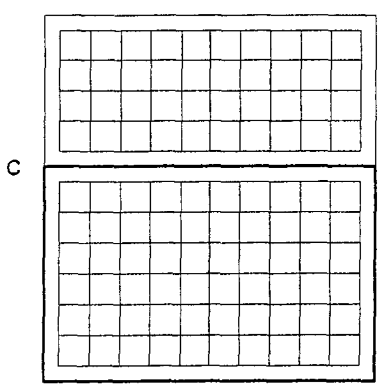
Annex C Detailed project plan



Annex D Cluster/Sample cost ratio

A cluster sampling survey has two different costs, the cost of each sample and the extra cost for each cluster as illustrated in Figure 9. The cost of sampling 1 BSU would be typically the salary and stationary cost for one sample while the extra cluster cost would be lodging transport, salary and other cost during hours that no surveying is done such as while travelling to a PSU.





For example imagine cluster A in Figure 9 which is one of 30 clusters in a given survey. The take is 20 BSU and the extra cluster cost is equivalent to cost of 60 sampled BSUs. This results in a cost ratio of 60/1=60 and a total cluster cost of sampling (30 cluster x 20 BSU) + (30 x 60 Cluster/Sample ratio) = <u>2400 BSUs</u>. If we want to double the sample size two possible ways of doing that is to double the number of clusters from 30 to 60 as shown in Figure 9B and keep the same cost per cluster. The survey would than cost the equivalent of sampling (60 cluster x 20 BSU) + (60 clusters x 60 Cluster/Sample ratio) = <u>4800 BSUs</u>. Another way of doubling

Figure 9: Cluster/Sample cost ratio

the sample size is doubling the take in each cluster and keeping the same number of clusters as shown in Figure 9C. The cost of the survey would than cost only the equivalent of sampling (30 cluster x 40 BSU) + (30 clusters X 60 Cluster/Sample ratio) = <u>3000 BSUs.</u>

Annex E Overview of factors considered for indicators

Indicators for hygiene practices

Personal hygiene

- Handwashing
- System for handwashing
- Handwashing demonstration as an 'indicator'
- Cleanliness of the interviewee/mother or caretaker
- Proper disposal of children's faeces
- Child uses diapers, underclothes/clean child's bottom.
- Bacteria on finger tips
- Cleanliness of sanitation facilities
- Toys or baby bottles on the ground

Food and water hygiene

- Water consumption
- Suitable water drawing mechanism for drinking water
- Storage conditions of food
- Cleanliness of dishes and utensils

Domestic and environmental hygiene

- Cleanliness of floor and compound surfaces
- Area in and around the household free of children's (human) and animals' faeces
- Waste disposal
- Flies and other vectors.
- Animals loose inside the house or the compound.

Indicators for access to improved sanitation

'Improved' sanitation

- Kind of use of the latrine
- Type of technology used
- Type of desludging or pit emptying technology used.
- Handwashing facilities near toilet.
- Groundwater pollution.

Definition of access

- Distance to the latrine.
- Proof of use by all household members
- Number of users per cubicle.
- Design life etc.
- Menstruation.
- Flies and other vectors.

Indicators for access to improved water sources

Improved drinking water sources

- Water sources used for drinking water
- Protected and unprotected sources.
- Ownership/level or access to drink water source.
- Water treatment.
- Water quality testing.

Access to improved water source

- Reliability of the water source.
- Time/Distance of water source
- Physical accessibility to the water
- Water quantity
- Amount of trips
- Energy needed to abstract and transport the water.

Indicators for access to non-drinking water

- Time/Distance of water source.
- Physical accessibility to the water
- Reliability of the water source.
- Water quantity
- Number of trips per day to collect water
- Energy needed to abstract and transport the water.
- Cost of water

Other aspects to access to improved water sources

- Seasonal variations
- Water storage
- Maintenance of water source
- Other activities related to water sources

Annex F Example from the final draft of the questionnaire

A1 is a question and A2 is an observation on hygiene behaviour while the decision model in flowchart format on the next page is on sanitation.

Question

A.1 Where do you usually wash your hands?

Conclusion

In dwelling/yard/plot	Needs more information. (go to question A2)
Somewhere else	In regards to this question it is presumed that this is **not** a *good hygiene practice* of the household in relation to handwashing.
Nowhere	

Origine

Question 33 of ORC Macro's DHS+ questionnaire.(ORC Macro 2001)

Rationale/assumptions

The question assumes that people are washing their hands as it only asks where they wash them. Handwashing is recognised as an important hygiene behaviour that limits faecal-oral transmission of pathogens. This question is part of a series of questions which aim to determine if handwashing is likely to be practised in the household, by examining whether the attributes essential for proper handwashing are available in the household. This assumes that the practice is unlikely to benefit health if it is only practised outside the dwelling, yard or plot.
The question does not discriminate between adequate ways and times of handwashing or whether all people in the household would adhere to this practice.

Remarks/limitations

An increase of hand washing will improve health in a household. This is because of its potential to reduce the number of pathogenic organisms on our hands. In many parts of the world hand washing is not perceived as related to health.(Hoque et al. 1995; Zeitlyn 1994) One person might use several methods of hand washing during the day, for example: rub the left hand with mud and rinse it with water after defecation; pour water over the right hand before eating; rub hands, arms legs and feet with water before prayer; or wash hands along with other parts of the body with soap in the course of a daily bath.(Hoque et al. 1995) Some actions like wetting hands before eating so the food does not stick to the hands (Kamanda 2002)can wrongly be associated with handwashing.
The question above only aims to identify households that are unlikely to practise hand washing (coded "2 & 3") as a way of reducing faecal-oral transmission. The households with code "1" will have the potential to practise 'proper' hand washing but additional questions are necessary.
The DHS manual (ORC Macro 2001)mentions that *"Washing hands, especially before handling food, can protect people from getting infected with various diseases such as diarrhoea"*. There are clear indications that in reference to the F-diagram first barrier handwashing (after defecation) is more important than second barrier handwashing such as before handling food (Curtis et al. 2000).

Observation

A.2 Ask to see the place and observe if the following items are present

Conclusion

When all items are present it is assumed that good hygiene practices are used in the household regarding handwashing. If any of the items are not present the opposite is assumed.

Origine

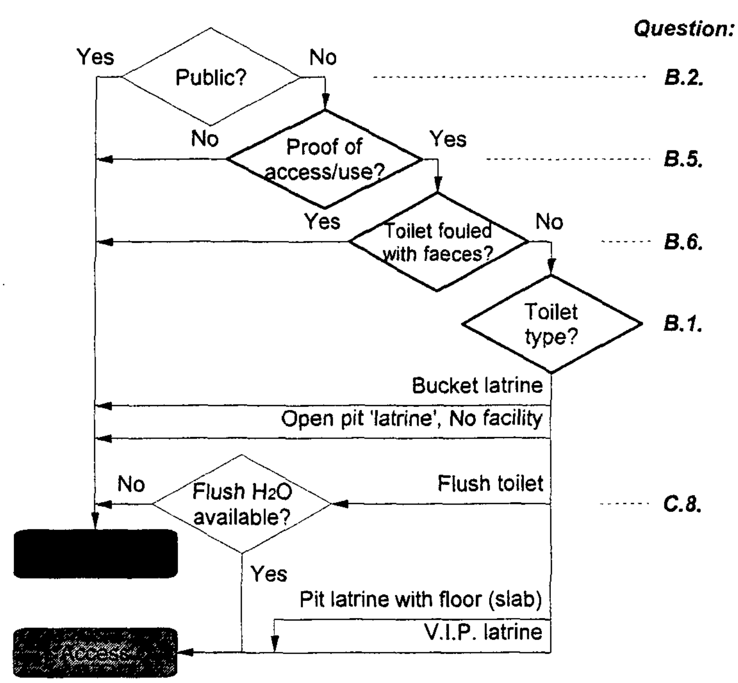
Question 34 of ORC Macro's DHS+ questionnaire.(ORC Macro 2001)

Rationale/assumptions

Handwashing as spot-observation is difficult to observe during the short period of the interviewer is in the household. An alternative which is easier to observe is whether attributes for proper handwashing are available in the household.

Remarks/limitations

We suggest to remove tap in answer 'A' to prevent the confusion if water or tap should be present. It is the water that should be present coming from a tap or not. We suggest adding bucket or sink to answer 'C'.
This question can only determine if it is possible for the household to practise handwashing by examining if attributes essential for appropriate handwashing are available in the household. Appropriate referring to handwashing as a way of reducing faecal-oral transmission of pathogens. The question excludes the households that are lacking essential attributes for hand washing. However the question does not allow us to determine if appropriate handwashing is practised by all members of the household in such a way and at times that it reduces faecal-oral transmission of pathogens.
An optional answer is suggested to determine the distance between the place/system for hand washing and the toilet if there is one



Annex G Screen shots from the sampling simulator

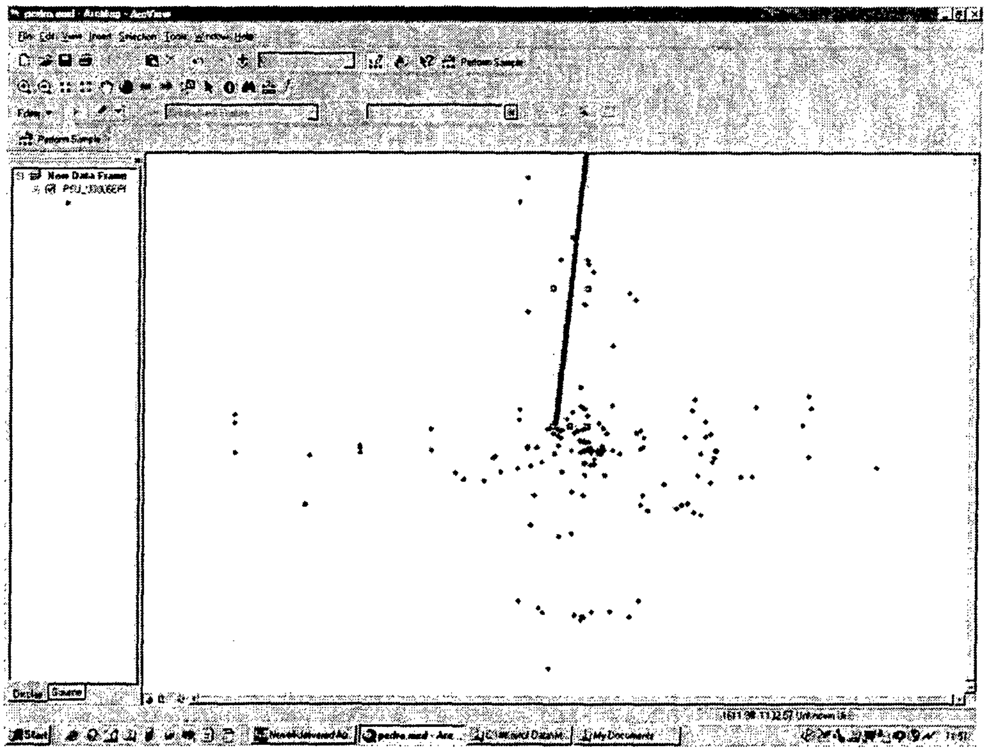


Figure 10: ArcGIS window with village and random sampling direction

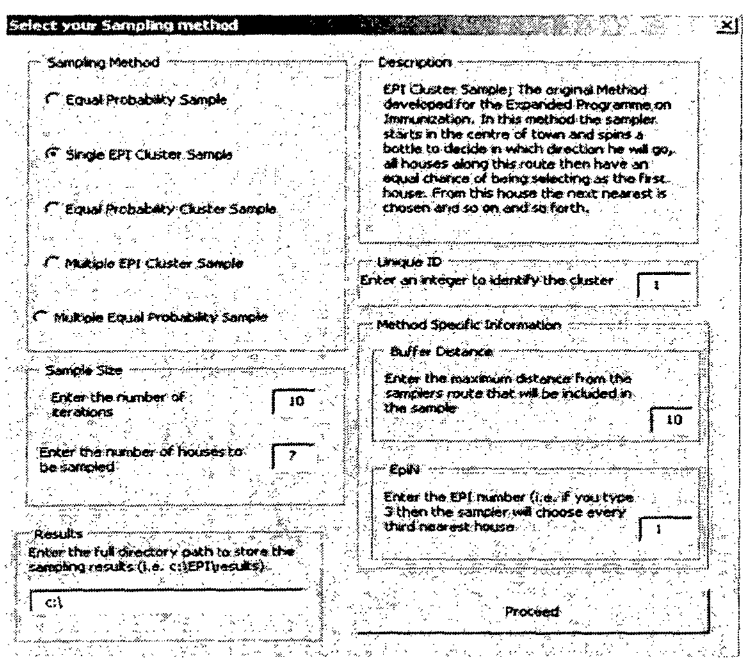


Figure 11: Screen dump of the window for selecting the sampling method.

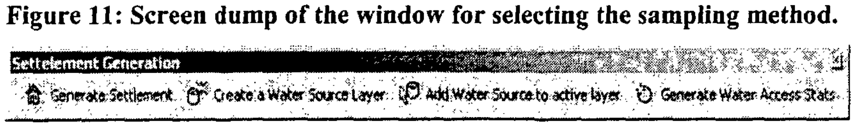


Figure 12: Toolbar for generating settlements and access data

Annex H The sampling simulation process

Figure 13 shows an overview of the various steps in the simulation process. The simulator is build in two different modules. The first 'creates' or 'completes' data with simulated data if no data or only partial data is available. If complete georeferenced data is available this step can be skipped as shown. The data used at this stage is stored in separate datasets for each PSU. The **simulator** can go trough the chosen PSU data set and will do typical 1000 iterations of the chosen sampling methodology. To add household density data on cluster and household level as well as the water access patterns and other information the data set has to be analysed by ArcMap™ 'Geostatistical Analyst' which is an existing programme in ArcGIS™. In the ArcMap™ 'Geoprocessor' which is another ArcGIS™ programme all the different clusters are merged in one data set. At this point there would be typically 1000 iterations of a full survey available. This data is converted with Stat/Transfer to a data set that can be read in Stata™. Data analyses in stat trough 'svy' (survey) commands and automatic creation of another Stata™ data set with the sampling iterations in the data set. From the original data set Anova analyses can be performed to analyse factors influencing the sampling accuracy. Steps between different programs require an operator to handle the data.

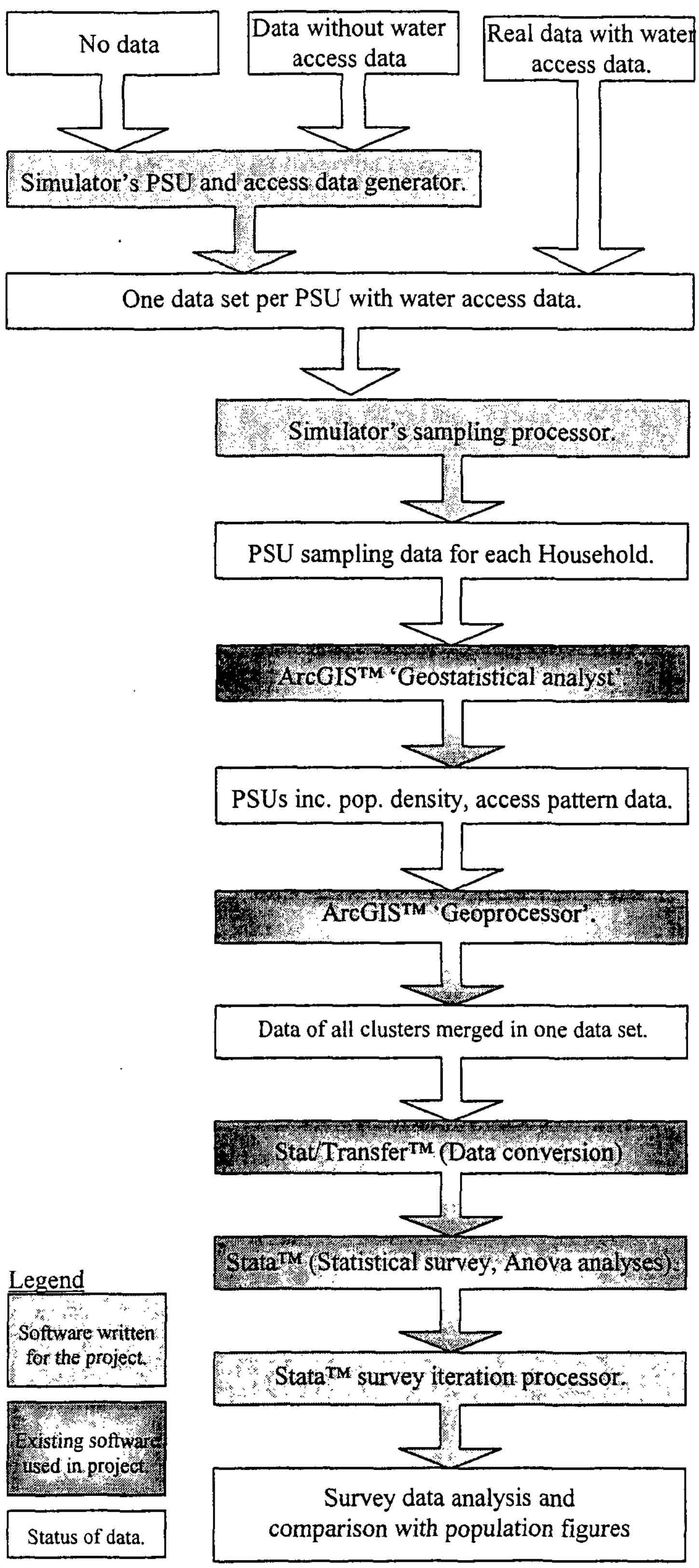


Figure 13: Sample simulation process.

Bibliography

Almedon, A., Blumenthal, U. and Manderson, L. (1997). Hygiene Evaluation Procedures, ODA, INFDC, LSHTM, UNICEF

Bennett, S. (1993). The EPI cluster sampling method: A critical appraisal. ISI 49th session, Firenze

Bennett, S., Radalowicz, A., Vella, V. and Tomkins, A. (1994). "A computer simulation of household sampling schemes for health surveys in developing countries." *International Journal of Epidemiology* 23(6): 1282-1291

Bennett, S., Woods, T., Liyanage, W. M. and Smith, D. L. (1991). "A simplified general method for cluster-sample surveys of health in developing countries." *World Health Statistics Quarterly* 44: 98-106

Boot, M. T. and Cairncross, S., Eds. (1993). Actions speak: The study of hygiene behaviour in water and sanitation projects. The Hague/London, IRC/LSHTM

Brogan, D., Flagg, E. W., Deming, M. and Waldman, R. (1994). "Increasing the accuracy of the expanded Program on Immunization's cluster survey design." *Annals of epidemiology* 4(4): 302-311

Cairncross, S. (2001). Minutes on WSSCC task force on monitoring 14/12/2001, Delft, WSSCC

Cairncross, S., Blumenthal, U., Kolsky, P., Moraes, L., *et al.* (1996). "The public and domestic domains in the transmission of disease." *Tropical Medicine and International Health* 1(1): 27-34

Curtis, V., Cairncross, S. and Yonli, R. (2000). "Domestic hygiene and diarrhoea - pinpointing the problem." *Trop Med Int Health* 5(1): 22-32

Curtis, V., Cousens, S., Mertens, T., Traore, E., *et al.* (1993). "Structured observations of hygiene behaviours in Burkina Faso: validity, Variability and utility." *Bulletin of the World Health Organization* 71(1): 23-32

DFID (1998). Guidance manual on water supply and sanitation programmes. Loughborough: WEDC & LSHTM, Department for International Development, Prepared by WELL

du V Florey, C. (1993). "Sample size for beginners." *BMJ* 306: 1181-4

Henderson, R. H., Davis, H., Eddins, D. L. and Foege, W. H. (1973). "Assessment of vaccination coverage, vaccination scar rates, and smallpox scarring in five areas of West Africa." *Bulletin of the World Health Organization* 48: 183-194

Henderson, R. H. and Sundaresan, T. (1982). "Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method." *Bulletin of the World Health Organization* 60: 253-260

Hoque, B. A., Mahalanabis, D., Alam, M. J. and Islam, M. S. (1995). "Post-defecation handwashing in Bangladesh: practice and efficiency perspectives." *Public Health* 109(1): 15-24.

Hunt, C. (2001). How safe is safe? A concise review of the health impacts of water supply, sanitation and hygiene. London, WELL(LSHTM / WEDC): 22

Kamanda, J. (2002). Personal communication. Bostoen K. London

Kendall, C. and Gittelsohn, J. (1994). Reliability and measures of hygiene behaviour: a case study. Studying hygiene behaviour: methods, issues and experiences. Cairncross S. and Kochar V., Eds. New Delhi, Sage

Kish, L. (1965). Survey sampling. New York ; Chichester, John Wiley and Son

Kolsky, P. and Blumenthal, U. (1995). "Environmental health indicators and sanitation-related disease in developing countries: limitations to the use of routine data sources." *World Health Statistics Quarterly* 48(2): 132-9

Kolsky, P. and Butler, D. (2002). "Performance indicators for urban storm drainage in developing countries (submitted for publications)." *Urban Water*

Lemeshow, S., Hosmer Jr, D. W., Klar, J. and Lwanga, S. K. (1990). Adequacy of sample size in health studies. Chichester, UK, John Wiley & Sons Ltd.

Lemeshow, S., Tserkovnyi, A. G., Tulloch, J. L., Dowd, J. E., *et al.* (1985). "A computer Simulation of the EPI Survey Strategy." *International Journal of Epidemiology* 14(3): 473-481

Levy, P. S. and Lemeshow, S. (1999). Sampling of Populations. New York, Wiley-Interscience

Macro (1996). Sampling manual. Calverton, Maryland USA, MACRO International

Manun'Ebo, M., Cousens, S., Haggerty, P., Kalengaie, M., *et al.* (1997). "Measuring hygiene practices: a comparison of questionnaires with direct observations in rural Zaire." *Trop Med Int Health* 2(11): 1015-21

Montanari, G. E. (1993). "Design effects and ratios of homogeneity in complex sampling designs." *Statistica* anno LIII(4): 633-46

ORC Macro (2001). Model "A" questionnaire with commentary for high contraceptive prevalence countries. Calverton, Maryland USA, ORC Macro

Pedersen, D. (1994). Qualitative and quantitative, two styles of viewing the world or two categories of reality. Studying hygiene behaviour; Methods used and experiences. Cairncross S. and Kochar V., Eds. New Delhi, Sage Publications: 72-83

Qaba, O. (1999). Intra-cluster Homogeneity in South African Survey Data. 52nd Session of the International Statistical Institute (ISI), Finland.http://www.stat.fi/isi99/proceedings/arkisto/varasto/qaba0988.pdf

Ruel, M. T. and Arimond, M. (2002). "Spot-check observational method for accessing hygiene practices: Review of experience and implications for programmes." *J Health Popul Nutr* 20(1): 65-76

Serfling, R. E. and Sherman, I. L. (1965). Attribute sampling methods for local health departments with special reference to immunization surveys. Atlanta, CDC Epidemiology program office

Silverman, D. (2000). Doing Quantitative Research, A practical handbook. London, Sage Publications

Thomsen, I., Tesfu, D. and Binder, D. A. (1986). "Estimation of Design Effects and Intraclass Correlations when using Outdated Measures of Size." *International Statistical Review* 54(3): 343-9

Turner, A. G., Magnani, R. J. and Shuaib, M. (1996). "A not quite as quick but much cleaner alternative to the expanded programme on the immunization (EPI) cluster survey design." *International Journal of Epidemiology* 25(1): 198-203

WHO (1992). Improving water and sanitation hygiene behaviours for the reduction of diarrhoeal disease. Geneva, WHO: 21

WHO and UNICEF (2000). Global Water Supply and Sanitation Assessment 2000 Report.http://www.unicef.org/programme/wes/pubs/global/global.htm

World Bank (1993). World Development Report 1993: Investing in Health. New York, Int. Bank for Reconstruction & Development

WSSCC (2000a). Iguaçu Action Programme. Fifth Global Forum on Water Supply and Sanitation, Foz do Iguaçu, WSSCC

WSSCC (2000b). Vision 21: A Shared Vision for Hygiene, Sanitation and Water Supply and a Framework for Action. Geneva, WSSCC: 64.www.wsscc.org/pdf/v21core.pdf

WSSCC (2002). Iguaçu Action Plan, WSSCC. **2002**.http:/www.wsscc.org/activities/iguacu/actprog.htm

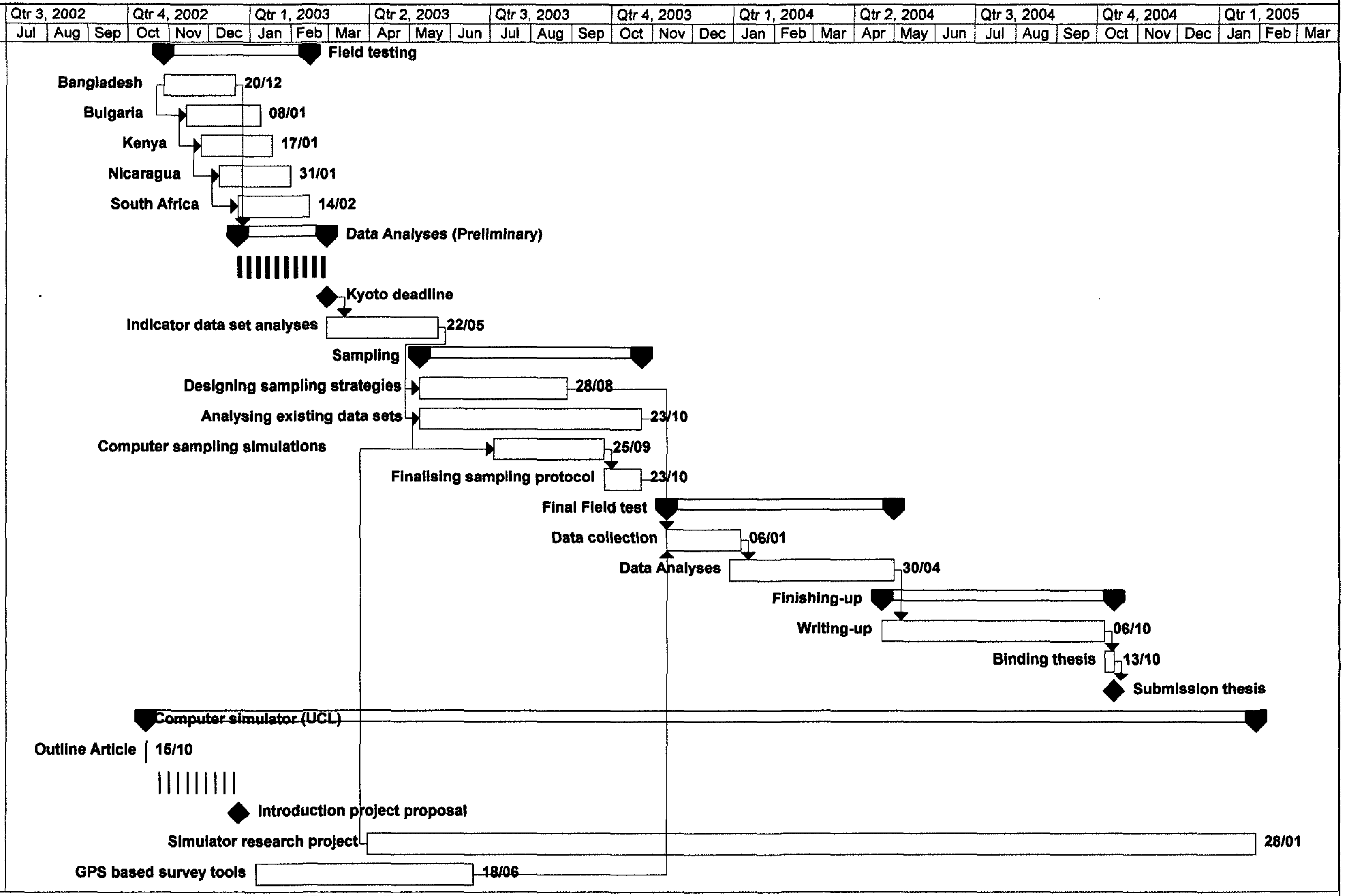
Yates, F. (1981). Sampling methods for censuses and surveys. London, Charles Griffin &Co Ltd

Zeitlyn, S. (1994). Measuring hygiene behaviour: The importance of meaning and definition. Studying Hygiene Behaviour: Methods, Issues and experiences. Cairncross S. and Kochar V., Eds. New Delhi, Sage Publications India: 49-58

Measuring Access to Water and Sanitation and application of Hygiene Practices
PhD research, Disease Control and Vector Biology Unit(DCVBU), Enivironmental Health Group.

Kristof BOSTOEN

ID	Name
1	**Field testing**
2	Bangladesh
3	Bulgaria
4	Kenya
5	Nicaragua
6	South Africa
7	**Data Analyses (Preliminary)**
8	**Analyses**
39	Kyoto deadline
40	Indicator data set analyses
41	**Sampling**
42	Designing sampling strategies
43	Analysing existing data sets
44	Computer sampling simulations
45	Finalising sampling protocol
46	**Final Field test**
47	Data collection
48	Data Analyses
49	***Finishing-up***
50	Writing-up
51	Binding thesis
52	Submission thesis
53	**Computer simulator (UCL)**
54	Outline Article
55	**Writing article**
65	Introduction project proposal
66	Simulator research project
67	GPS based survey tools



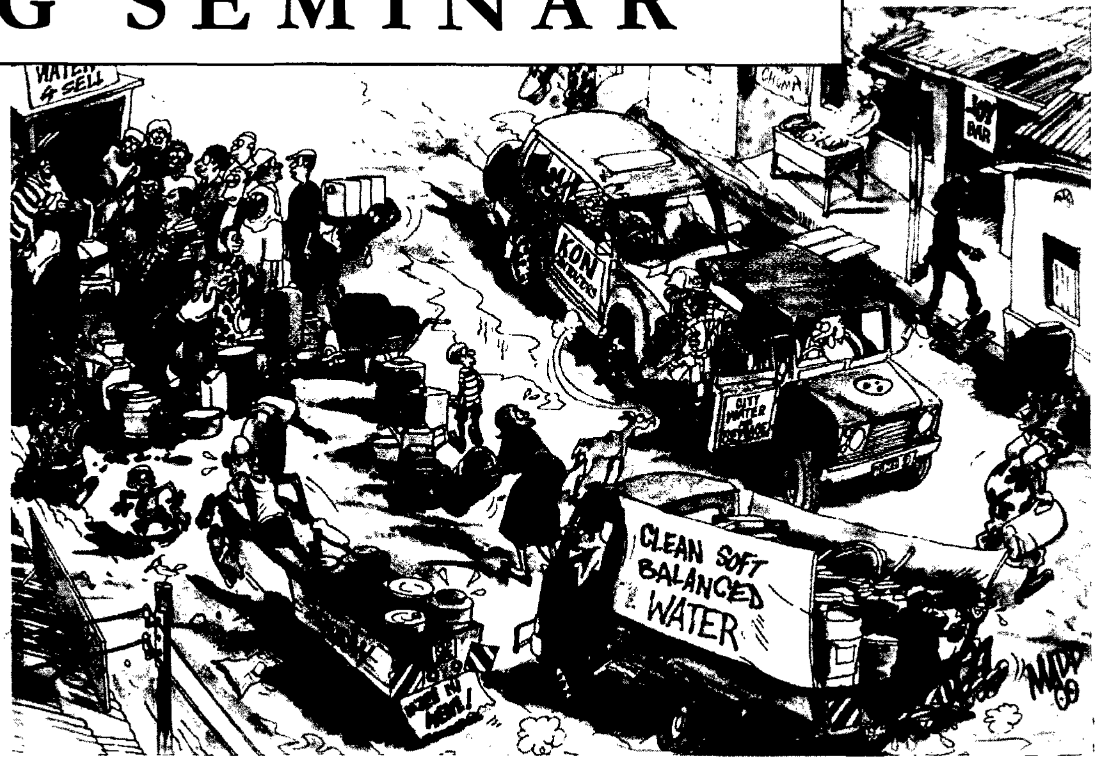
Project: Measuring Access & Practice
Date: Fri 11/10/02

Task	Summary	Rolled Up Progress	Group By Summary
Critical Task	Rolled Up Task	Split	
Progress	Rolled Up Critical Task	External Tasks	
Milestone	Rolled Up Milestone	Project Summary	

DCVB UNIT

Wednesday 30th Oct 2002
Room 364
12.45pm till 2.00pm

UPGRADING SEMINAR



Kristof Bostoen

Measuring Access and Practice (Water, Sanitation and Hygiene)

ALL WELCOME