

Zambia - World Health Survey 2003

World Health Organization (WHO)

Report generated on: February 18, 2014

Visit our data catalog at: <http://apps.who.int/healthinfo/systems/surveydata/index.php>

Sampling

Sampling Procedure

SAMPLING GUIDELINES FOR WHS

Surveys in the WHS program must employ a probability sampling design. This means that every single individual in the sampling frame has a known and non-zero chance of being selected into the survey sample. While a Single Stage Random Sample is ideal if feasible, it is recognized that most sites will carry out Multi-stage Cluster Sampling.

The WHS sampling frame should cover 100% of the eligible population in the surveyed country. This means that every eligible person in the country has a chance of being included in the survey sample. It also means that particular ethnic groups or geographical areas may not be excluded from the sampling frame.

The sample size of the WHS in each country is 5000 persons (exceptions considered on a by-country basis). An adequate number of persons must be drawn from the sampling frame to account for an estimated amount of non-response (refusal to participate, empty houses etc.). The highest estimate of potential non-response and empty households should be used to ensure that the desired sample size is reached at the end of the survey period. This is very important because if, at the end of data collection, the required sample size of 5000 has not been reached additional persons must be selected randomly into the survey sample from the sampling frame. This is both costly and technically complicated (if this situation is to occur, consult WHO sampling experts for assistance), and best avoided by proper planning before data collection begins.

All steps of sampling, including justification for stratification, cluster sizes, probabilities of selection, weights at each stage of selection, and the computer program used for randomization must be communicated to WHO

STRATIFICATION

Stratification is the process by which the population is divided into subgroups. Sampling will then be conducted separately in each subgroup. Strata or subgroups are chosen because evidence is available that they are related to the outcome (e.g. health, responsiveness, mortality, coverage etc.). The strata chosen will vary by country and reflect local conditions. Some examples of factors that can be stratified on are geography (e.g. North, Central, South), level of urbanization (e.g. urban, rural), socio-economic zones, provinces (especially if health administration is primarily under the jurisdiction of provincial authorities), or presence of health facility in area. Strata to be used must be identified by each country and the reasons for selection explicitly justified.

Stratification is strongly recommended at the first stage of sampling. Once the strata have been chosen and justified, all stages of selection will be conducted separately in each stratum. We recommend stratifying on 3-5 factors. It is optimum to have half as many strata (note the difference between stratifying variables, which may be such variables as gender, socio-economic status, province/region etc. and strata, which are the combination of variable categories, for example Male, High socio-economic status, Xingtao Province would be a stratum).

Strata should be as homogenous as possible within and as heterogeneous as possible between. This means that strata should be formulated in such a way that individuals belonging to a stratum should be as similar to each other with respect to key variables as possible and as different as possible from individuals belonging to a different stratum. This maximises the efficiency of stratification in reducing sampling variance.

MULTI-STAGE CLUSTER SELECTION

A cluster is a naturally occurring unit or grouping within the population (e.g. enumeration areas, cities, universities, provinces, hospitals etc.); it is a unit for which the administrative level has clear, nonoverlapping boundaries. Cluster sampling is useful because it avoids having to compile exhaustive lists of every single person in the population. Clusters should be as heterogeneous as possible within and as homogenous as possible between (note that this is the opposite criterion as that for strata). Clusters should be as small as possible (i.e. large administrative units such as Provinces or States are not good clusters) but not so small as to be homogenous.

In cluster sampling, a number of clusters are randomly selected from a list of clusters. Then, either all members of the chosen cluster or a random selection from among them are included in the sample. Multistage sampling is an extension of cluster sampling where a hierarchy of clusters are chosen going from larger to smaller.

In order to carry out multi-stage sampling, one needs to know only the population sizes of the sampling units. For the

smallest sampling unit above the elementary unit however, a complete list of all elementary units (households) is needed; in order to be able to randomly select among all households in the TSU, a list of all those households is required. This information may be available from the most recent population census. If the last census was >3 years ago or the information furnished by it was of poor quality or unreliable, the survey staff will have the task of enumerating all households in the smallest randomly selected sampling unit. It is very important to budget for this step if it is necessary and ensure that all households are properly enumerated in order that a representative sample is obtained.

It is always best to have as many clusters in the PSU as possible. The reason for this is that the fewer the number of respondents in each PSU, the lower will be the clustering effect which increases sample variance and effectively reduces our estimating power. WHO requires an absolute maximum of 50 respondents per PSU, and ideally would suggest 20-30. This means that for a sample size of 5000 respondents, 100- 200 PSU clusters should be taken into the sample. Calculating that, roughly, one fifth of the total number of PSU clusters in a country will be randomly selected into the survey sample, the sampling frame should consist of 500-1000 PSU clusters.

PROBABILITY SAMPLING

Probability sampling means that every single individual in the sampling frame has a known and non-zero chance of being selected into the survey sample. Non-probability methods of sampling such as quota or convenience sampling and random walk, may introduce bias into the survey, will throw survey findings into question, and are not accepted by WHO.

The probability of selection into the survey sample for each cluster will be proportional to its relative size. Systematic Sampling Systematic sampling is the ordered sampling at fixed intervals from a list, starting from a randomly chosen point. Typically, systematic sampling is not used at the first stage of sampling (selection of PSUs) because it renders the estimation of sampling error difficult.

Systematic sampling is recommended at the SSU, TSU, and household selection stages of sampling. Systematic sampling may be linear or circular.

SELECTION OF HOUSEHOLDS

The Household is a device used to get at the individual. The household is the sampling unit while the individual is the observational unit. While it would be preferable to randomly select from a list of all eligible persons in a country, such lists, with a few exceptions, are not available, so we must employ a final cluster, the household, to get at our observational units.

Households will be selected from lists of dwelling units. Non-probabilistic methods of household selection such as the random walk are not acceptable. Such lists are typically available from population registries, household listings, voter lists and census list. As it is essential to include all households in the sampling frame, an assessment of the methodology employed to select households must be made:

- How much has the population changed since these lists were made?
- Completeness of coverage. Are there unregistered populations (e.g. slums)
- Population shifts
- Changes in Registry

QUALITY

Almost all lists will suffer from routine problems. WHO recommends that survey institutions manually enumerate all the households in the sampling units randomly selected into the survey sample. If existing lists or registries will be used, then a detailed analysis of their quality must be made and they must be updated to ensure that there is no exclusion of households from the survey sampling frame.

SELECTION OF INDIVIDUALS FROM HOUSEHOLD ROSTER

All members of each household selected into the survey sample will be enumerated on the household roster. A member of the household is defined as someone who usually stays in the household, sleeps and shares meals, who has that address as primary place of residence, or who spends more than 6 months a year living there. Country-specific variations in this standard definition are allowed in consultation with WHO.

KISH TABLES

The respondent for the survey will be selected among all eligible members of the household using Kish tables. Kish tables provide a method by which each eligible person in a household has an equal probability of selection into the survey sample. It is extremely important for the representativeness of the survey sample and the integrity of the survey that the Kish tables

are properly implemented. All interviews where the Kish selection method is not properly implemented will be rejected. The Kish technique allows adequate representation for all the persons in the household.

Response Rate

The proper and complete enumeration and description of the entire household is a critical component of the survey process. The household roster must be completed for all households selected randomly into the survey sample, whether they agreed to participate in the survey or not. It is only in this way that we can collect the basic information required to estimate the non-response bias in the survey.

The requirement of augmenting the survey sample size to adjust for estimated non-response is necessary to ensure that we have adequate persons in the sample to have the power to make precise estimates. This does not, however, account for the bias that is created by non-response, since non-responders are often different from responders with respect to key variables that are linked to the domains under study in the survey. All effort, therefore, must be made to minimise non-response, and to interview as many people in the survey sample as possible. A detailed discussion of refusal conversion methods, survey awareness raising, and call-backs is found in the WHS Survey Manual.

There are two possible scenarios of non-response:

1) The interviewer completes the household roster and the randomly chosen respondent refuses to participate.

2) The interviewer is refused access to the household and is unable to fill in the household roster.

In second scenario, sites must ensure that, at least, pages 00.1 and 00.3 of the Coversheet are completed for the household. In addition, if available from census information, the number of adult (18 years of age or older) males and females in the household, and their respective ages should be provided. It is important to note that the completion of the household roster serves a purpose above and beyond providing a list from which a respondent will be selected.

The demographic and other information collected in the household roster and requested from sites serves to calculate the denominators for statistical analysis of the survey data; without the information in the household roster, we would not be able to determine the health-related outcome rates in your country.

Weighting

Countries has provided WHO with the population sizes, probabilities of selection and sampling weights of all sampling units for each stage of the sampling process. Since clusters are often of unequal size, sampling weights are necessary to be able to reconstruct population estimates from our sample estimates.

The weights essentially describe the number of persons in the sampling frame represented by each person in the cluster (i.e. each person in County 1 represents 12.5 people, each person in County 2 represents 9.1 persons etc.). Weights for SSUs, TSUs, etc. are calculated in the same way. The probability of selection of the elementary unit, the household, is not proportional to the number of people in the household. Rather, the household level weights will be generated at the time of respondent selection within the household. The number of households selected within each chosen sampling unit will be proportional to the total number of households in that sampling unit. All households in each unit will have equal probability of selection.

Questionnaires

No content available

Data Collection

Data Collection Dates

Start	End	Cycle
2003	2003	N/A

Data Collection Mode

Face-to-face [f2f]

Data Collection Notes

Participation in the survey is voluntary and the respondent can refuse to be interviewed. The interviewer is responsible for explaining what the survey is about, providing all the necessary information, and making sure the respondent understands the implications of his / her participation before giving his / her consent. The information given should be simple and clear and adapted to the respondent's level of understanding.

Consents must be documented by asking the respondents to sign an Informed Consent Forms (Household Informant Consent Form; Individual Consent Form) before doing the interview. These forms must mention who will be doing the study, the types of questions that will be asked, why the study is being done, and who will have access to the information provided. The interviewer must check that the respondent has read and understood the form before signing, and should offer to go over it with him /her emphasizing the different items mentioned. If the respondent is illiterate or unable to read for himself / herself (e.g. due to a visual impairment), the form will be read and explained to him / her. In cases where it is not appropriate for the respondent to sign the form, the interviewer alone will sign the form.

In cases where the respondent is being dissuaded from, or coerced into, participating in the study by a third party such as a spouse, relative or any other member in the community, the interviewer should make it clear that it is the respondent alone who must decide whether or not s/he wishes to be interviewed.

Supervision

FIELD SUPERVISION

The role of the supervisor is to monitor the progress and quality of the data collection, and ensure that interviewers are performing work. This includes handling the logistics of the survey, coordinating with other staff, recruiting and training additional interviewers, and supervision.

Supervision must be given to interviewers before, during, and after the interview. Supervisors must not only check that contact procedures are followed correctly but that interviews are conducted appropriately. They should be present for 10 interviews and ensure that standardized interviewing techniques are observed when asking questions, clarifying, probing and giving feedback in a non-directive manner. After the interview has been conducted, they must also check that data is coded and entered correctly. Supervisors must give feedback and debrief on a regular basis, in order to update the organization responsible for the study on the progress of the survey and any problems that have arisen.

Data Processing

No content available

Data Appraisal

No content available