

Part 2: Planning and Set Up

Overview

In this Part This Part covers the following topics

Topic	See Page
Section 1: Planning and Preparing a STEPS Survey	2-1-1
Section 2: Preparing the Sample	2-2-1
Section 3: Preparing a STEPS Site	2-3-1
Section 4: Preparing the Data Entry Environment	2-4-1
Section 5: Preparing the Data Analysis Environment	2-5-1

Section 1: Planning and Preparing a STEPS Survey







Overview

Introduction This section covers the tasks that need to be conducted to plan for your STEPS chronic disease risk factor survey.

Intended audience This section is primarily designed to be used by those fulfilling the following roles:

- STEPS Site Coordinator
 - STEPS Coordinating Committee.
-

Tasks and timeframes The chart below shows the main tasks and indicative timelines covered in this section.

Task Name	Duration	Month 1	Month 2
Develop implementation plan	1 week		
Identify scope of STEPS survey	1 week		
Gain ethical approval	1 week		
Schedule data collection	2 days		
Adapting and translating the STEPS Instrument	1 week		
Pilot test	1 week		

In this section This section covers the following topics:

Topic	See Page
The STEPS Implementation Plan	2-1-2
Identifying the Scope of the STEPS Survey	2-1-5
Choosing a Chemistry Screening Method for Step 3	2-1-9
Applying for Ethical Approval	2-1-10
Timeframes and Data Collection Considerations	2-1-11
Number of Staff Required	2-1-13
Scheduling Data Collection	2-1-14
Adapting the STEPS Instrument	2-1-15
Translating STEPS Documents	2-1-19
Pilot Testing	2-1-21

The STEPS Implementation Plan

Introduction You will need to create a detailed STEPS implementation plan for all stakeholders involved in the surveillance process.

Purpose The purpose of the implementation plan is to:

- outline the scope of the surveillance and desired goals
- identify required resources
- create an action plan
- develop a communication strategy
- provide a well-planned budget as a basis for funding.

Requirement The content of the implementation plan should be developed using the guidelines in the sections below. Once complete, it should be agreed upon by the Coordinating Committee after wide consultation and discussion, and sent to the WHO Geneva STEPS team for review.

Core topics The topics that should be covered in the implementation plan and references to appropriate sections in the manual where guidelines can be found are listed in the table below:

Topics	Detail	Reference
Executive Summary	Provide high level summary of main points including: <ul style="list-style-type: none">• current situation• goals• scope• resources• budget.	
Current Situation	Specify: <ul style="list-style-type: none">• if a risk factor survey has already been conducted in this setting;• the availability of risk factor data in this setting;• if there is an infrastructure (human capacity, equipment, etc.) on which STEPS could be built;• the rationale for conducting chronic disease risk factor surveillance.	Part 1, Section 1

Continued on next page

The STEPS Implementation Plan, Continued

Core topics (cont.)

Topics	Detail	Reference
Goals and Objectives	<ul style="list-style-type: none"> • Identify planned goals and the use of collected information to: <ul style="list-style-type: none"> – describe the current levels of chronic disease risk factors in this population; – track the direction and magnitude of risk factor trends; – plan or evaluate a health promotion or preventive campaign; – collect data from which to predict likely future demands for health services. • Specify objectives that support gathering 'essential' information only. • Describe broad timeframes. 	Part 2, Section 1
Scope	<ul style="list-style-type: none"> • Specify the scope of surveillance to be conducted (Step 1, Step 2 and Step 3, plus coverage of core, expanded and optional items). • Specify if future STEPS surveillance can be assured. 	Part 2, Section 1
Sampling methods	<ul style="list-style-type: none"> • Identify the sample size and sample frame that will be used.* • Identify geographical coverage. • Describe sampling design. 	Part 2, Section 2
Resources	<ul style="list-style-type: none"> • Specify required resources in terms of all personnel and equipment required for STEPS surveillance. • Describe resources that have already been committed or which are expected, including support from WHO. • Specify resources expected from other organizations involved. 	
Action Plan	Provide a chart of the main tasks, with estimated start dates and timeframes for completion of each phase.	
Communication strategy and publicity	Specify methods for informing and involving community leaders, members of the public, and the media in the STEPS surveillance project to gain commitment and support.	
Reporting and Disseminating Results	Describe to whom and how the results will be reported and disseminated.	Part 4, Section 4

Continued on next page

The STEPS Implementation Plan, Continued

Core topics (cont.)

Topics	Detail	Reference
Budget	Provide a detailed budget that includes: <ul style="list-style-type: none">• total funds required for each year planned to implement all STEPS activities as identified in the Scope (including future surveys);• source of funds;• funding gap.	

* During the planning phase of the survey, it is fundamental to determine the size of the sample as this will impact operational considerations, such as the number of interviewers required. There will have to be a compromise in which the precision requirements of the estimates are weighted against various constraints such as available budget, resources and time.

Implementation plan template

A STEPS Implementation Plan Template can be found in Part 6, Section 1.

Identifying the Scope of the STEPS Survey

Introduction

To develop a STEPS implementation plan, the scope of the STEPS Instrument being covered must be clearly defined.

The WHO STEPwise Instrument

The focus of the WHO STEPwise approach to surveillance of chronic disease risk factors is reflected in the core modules of the STEPS Instrument.

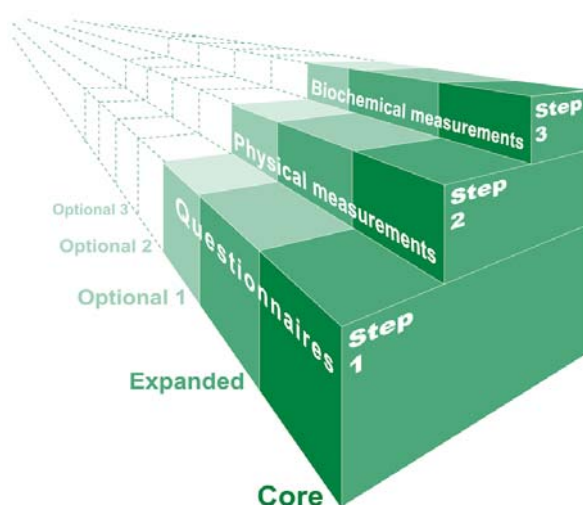
Step 1 core and expanded information will provide basic data on behavioural risk factors. Including Step 2 core and expanded physical measurements will provide useful additional data on excess body fat, raised blood pressure and heart rate.

Including Step 3 biochemical measurements is recommended only in countries that are well resourced, and will provide data on raised blood glucose and cholesterol levels.

Note: The STEPS Instrument can be found in Part 5, Section 1.

STEPS diagram

The diagram below shows each of the Steps.



Note: For guidance on implementing each of the Steps, please refer to the pages that follow. To fully understand each item covered in the STEPS Instrument, please see the Question-by-Question Guide in Part 5, Section 2.

Continued on next page

Identifying the Scope of the STEPS Survey, Continued

Step 1 core questions

All countries should undertake the Step 1 core items. This is an affordable option that will provide basic demographic information and measures of:

- tobacco smoking
 - alcohol consumption
 - fruit and vegetable consumption
 - physical activity.
-

Step 1 expanded questions

Countries should undertake Step 1 expanded to:

- describe demographic breakdowns (e.g., ethnicity and employment status);
- collect information on ex-smokers and smokeless tobacco (if it is used in your country);
- capture information on drinking with meals and drinking in the past 7 days;
- collect information about oil and fat consumption and meals outside a home;
- capture sedentary behavior;
- describe blood pressure history;
- describe diabetes history.

This level of detail is recommended for most countries/sites.

Step 2 core

Most countries/sites should undertake the Step 2 core items. These are affordable and can be done at the same time as Step 1, using the same data collection staff. Step 2 core will provide measures of:

- height and weight
 - waist circumference
 - blood pressure.
-

Step 2 expanded

Countries should undertake Step 2 expanded only if they need to know more about obesity and physical fitness. Step 2 expanded will provide measures of:

- hip circumference
 - heart rate.
-

Continued on next page

Identifying the Scope of the STEPS Survey, Continued

Step 3 core Sites should undertake Step 3 core items only if they are well resourced and have a need to detect the prevalence of diabetes and raised cholesterol. Step 3 core will provide measures of:

- blood glucose
- total cholesterol.

Note: For most countries, the cost of this option makes it not viable to survey all participants. One useful alternative is to conduct Step 3 tests on a sub-sample of the participants.

Step 3 expanded Sites should undertake Step 3 expanded only if they need to know about abnormal lipid profiles as a risk factor for cardiovascular diseases. Step 3 expanded will provide measures of:

- triglycerides
 - HDL cholesterol.
-

Optional questions Some sites may wish to go beyond Step 1 and Step 2 core and expanded to describe the prevalence of other specific health problems.

This may be achieved by asking the additional 'optional' questions in Step 1, and taking additional 'optional' measurements in Step 2.

Step 1 and 2 optional If you want to capture the prevalence of a particular health problem, you can add optional items to Step 1 and Step 2. For example:

If you need to	Then add
Assess a particular health problem, such as prevalence of: <ul style="list-style-type: none">• injuries and violence• mental health issues• oral health issues.	Optional questions to Step 1.
Conduct physical measurements of a particular health problem, such as oral health.	Optional measurements to Step 2.
Link the STEPS survey to other population surveys.	Appropriate optional questions.

Continued on next page

Identifying the Scope of the STEPS Survey, Continued

Considerations When countries add additional questions to Step 1 and Step 2 to tailor the Instrument to a local context, the cost of collection, analysis and presentation of the information escalates.

Adding more questions and local information also adds to the burden on participants in the surveys, and thus threatens the level of participation in future surveys in the same population.

Step 3 core generally doubles the cost of the survey.

Note: Data checking and cleaning have been estimated to account for about 20% of the total cost of population surveys.

Choosing a Chemistry Screening Method for Step 3

Introduction Blood chemistry screening methods are widely used in community-based screening programs and public health surveillance for measurements of:

- glucose
- cholesterol
- triglycerides
- high density lipoproteins (HDL).

Note: This section applies only to those countries undertaking Step 3.

Dry or wet chemistry? Decide whether dry (blood collection from the fingertip) or wet ('gold standard', laboratory-based drawing of blood samples) chemistry will be used.

Staff, training and clinic equipment will be dependent on the choice.

The table below lists the advantages and disadvantages of both dry and wet chemistry.

Type	Advantages	Disadvantages
Dry	<ul style="list-style-type: none"> • rapid results available on-site • small sample volumes • no sample transport required • no pre-analytical variables • convenient to participants • viable option for less-resourced and unstable settings. 	<ul style="list-style-type: none"> • operators need good training and supervision • less accurate results as compared to wet methods
Wet	<ul style="list-style-type: none"> • accurate results • centralized laboratory with trained staff and good internal and external quality control • preferred method for well-resourced settings 	<ul style="list-style-type: none"> • more costly than dry methods

Devices for dry chemistry The table below lists a selection of dry chemistry devices, along with information on which of the Step 3 measurements they perform. Please note that this selection is just a list of examples. For more information on these and other devices, please contact the WHO Geneva STEPS team.

Device	Measurement
Reflotron Plus	Blood glucose, total cholesterol, triglycerides and HDL cholesterol (86)
Accutrend Plus	Blood glucose, total cholesterol, triglycerides and HDL cholesterol (87)
HemoCue 201 DM	Blood glucose (88)
Accu-Check	Blood glucose (89)
Cholestech LDX	Total cholesterol, triglycerides and HDL cholesterol (90)

Applying for Ethical Approval

Introduction

Every STEPS survey proposal should undergo technical and ethical review and approval. This is to ensure that the STEPS survey:

- is conducted in a technically and ethically sound manner
 - recognizes and protects the rights of participants
 - obtains access to information used in the sampling frame.
-

Process

Ideally, ethical approval should be sought by submission of a proposal and application to a national ethics review committee or other relevant body.

Where no such established process exists, it is recommended that an application for ethical review be prepared and submitted through an ad hoc local mechanism within the Ministry of Health.

Informed consent

Informed consent needs to be obtained from every survey participant before conducting the interviews. See Part 4, Section 1 for more details on gaining informed consent.

Making a submission

Follow the steps below to make a submission.

Step	Action
1	Determine if the ethics committee has a template for proposals which they require researchers to use.
2	Draft a formal submission (See Part 6, Section 1 for guidance on what to include in an ethical clearance submission).
3	Identify and contact the relevant committees, seeking guidance on rules, submission processes and procedures, and committee sitting times.
4	Adapt submission as necessary and submit to the appropriate committee, requesting guidance on expected timeframe for approval.
5	Follow up with the committee to gain clearance.

Note: The STEPS regional focal points and the WHO Geneva STEPS team can provide further advice on making a submission.

Expected timeframes

Preparing and obtaining approval for submissions to ethics committees can take weeks and even months depending on their rules of operation in the site and how often the committees sit.

Timeframes and Data Collection Considerations

Introduction

Data collection should be carefully planned to take place over a defined period of time and within appropriate seasons.

General timeframes

The table below provides a guide to estimated timeframes for each phase in a STEPS survey.

Phase	Suggested timeframes
Planning and scoping	1-2 weeks
Recruitment and training	3-4 weeks
Data collection	8-10 weeks
Data entry	4-6 weeks
Data analysis and reporting	2-4 weeks

Data collection

If possible, you should aim to complete data collection within a period of eight to twelve weeks.

Some key factors to consider when identifying an appropriate time to conduct the survey include:

Factors to consider	Guidelines
Seasons	<ul style="list-style-type: none">• Confine the survey period to one season to avoid dietary changes.• Avoid festive seasons (Ramadan, Christmas, and other national holidays).• Avoid rainy seasons where it may be physically difficult to get to individual households.• Avoid seasons when food is in unusually short supply.
Calendar year	Confine the survey period to one calendar year.
Major events	Avoid data collection during periods prior to local, regional or national elections to avoid confusion with political campaigners.
Civil unrest, turmoil, famine, etc.	It is not appropriate to conduct STEPS during times when more pressing matters occupy the minds and lives of the population. Sometimes it may be necessary to defer or cease a STEPS survey because of an intervening event.
Collection timeframe	Keep timeframe as close as possible (within reason) to the recommended timeframe.

Continued on next page

Timeframes and Data Collection Considerations, Continued

Data collection locations

It is recommended that both Step 1 and Step 2 are conducted in household settings.

Step 3 should be conducted in a clinic setting. This is recommended for:

- hygiene standards when taking blood samples
 - quality control
 - more accurate results.
-

Number of Staff Required

Data collection staff Use the following table as a guide to help determine the number of data collection staff required to interview. A final sample size of 4,500 participants was used to determine the numbers in the table below. The numbers may need to be adjusted depending on the final sample size of your survey.

Option	If you conduct	Average number of interviews	Number of interviewers	Number of supervisors
1	• Step 1 core and expanded	6-7	16-20	2-4
2	• Step 1 core and expanded • Step 2 core	4-6	20-24	4-6
3	• Step 1 core and expanded • Step 2 core and expanded	4-6	24-32	6-8
4	• Step 1 core and expanded • Step 2 core and expanded • Step 3 core (and expanded)	4-6	24-32	6-8

Note:

- The average number of interviews represents the number of interviews conducted or measurements taken by one interviewer during an eight hour working day.
- If you increase the size of the survey beyond 4,500, or extend this timeframe for data collection, these indicative numbers would change accordingly.

Data collection teams Consider the following factors when putting together interview teams:

- Consider allocating between two and four interviewers per team, each assigned to different areas.
- In some sites, you may wish to pair male and female interviewers.
- One supervisor should be responsible for between two and five teams.

Data entry staff Use the following table to help determine the number of data entry staff required to enter the completed instruments, twice, within an optimal four week period.

Total instruments received per day	Average instruments entered per staff per day	Number of data entry staff	Number of Supervisors
80-100	15-30 (depends on length of site-specific instrument)	8-12	1

Scheduling Data Collection

Introduction

To ensure data collection is completed within the planned 8 to 12 week timeframe, you will need to carefully schedule interviews.

When to schedule data collection

Ideally, as soon as your implementation plan and funding have been approved, the STEPS materials have been translated, and the sample has been drawn, participant lists should be collated and data collection scheduled.

Considering the size of this task, however, in practical terms it is recommended that data collection is conducted after the recruitment and training of data collection staff. This way trained interviewers can be used to compile the lists and establish contact with individual households.

Step 1 and Step 2 household settings

In some settings, evenings and weekends are generally preferred for interviewing, especially in urban areas.

This needs to be adapted on an individual country basis, as weekends in some countries are not the same days as in others.

Step 3 clinic setting

Schedule participants for blood collection into early morning slots at the clinic. This is because of the fasting requirement.

Adapting the STEPS Instrument

Introduction Use of a standardized STEPS Instrument enables comparisons both within the country over time and also between countries. However, the degree to which the Instrument can be standardized across cultures or settings can be limited.

When to adapt the Instrument Adaptations may need to be made to the STEPS Instrument to provide valid data for the surveillance site or to address the needs for information on other risk factors.

The following table provides guidance on when the Instrument can be adapted to local requirements.

Item	If...	Then...	Notes
Terminology	The terms used in some core questions do not fit the cultural setting (e.g. occupations).	Alter the term for local relevance, but ensure the original meaning is retained.	Changing the wording can easily alter the meaning of a question. Seek advice before changing questions.
Additional information	You require additional data on risk (e.g. exposure to indoor smoke) and you have available resources.	Add selective, but limited questions as expanded or optional items.	Inserting them in the middle of the core/expanded sections may alter the meaning of the questions. Insert them where they best fit so that they work with the flow of the other questions.
Link to previous data	You require specific data to link to previous surveys.	Add selective, but limited questions as expanded or optional items.	Insert the questions where they best fit so that they work with the flow of the other questions.
Questions not applicable	Questions asking about a particular health behaviour which are not applicable in your setting, (e.g. alcohol, or smokeless tobacco).	Drop these questions.	Look first at the fact sheet analysis guide and data book to see the impact on removing questions on the analysis.
Presentation	You want the skip instructions to correspond to the question numbers.	Change the skip instructions from the code identifier to the question number.	Only change the skip patterns from the codes to the question number once the questionnaire is finalized.

Continued on next page

Adapting the STEPS Instrument, Continued

Rules

The list below provides some fundamental rules that must be observed when tailoring the STEPS Instrument to create your site-specific instrument.

- Never delete a question or measure from the core sections (unless question is not applicable in your setting).
 - Never change the standard coding numbers.
 - Place additional questions or measures where they best fit within relevant section as an expanded or optional item.
 - Do not place additional questions or measures in between core questions or measures.
 - Code all added questions or measures with the letter 'X' plus a number (e.g. X1, X2...).
 - Remove from the Instrument the expanded sections and Steps (i.e. 2 and/or 3) that are not being covered by your site.
 - Amend the skip instructions if expanded or optional items are added to the any section.
 - Review all skip instructions.
 - Send a draft of your tailored STEPS Instrument to the WHO Geneva STEPS team for review before finalizing.
-

Process

The process of adapting the STEPS Instrument involves the following key stages:

Stage	Description
1	Identifying questions that require local adaptation.
2	Adapting wording or adding questions and adjusting skip instructions.
3	Adapting other forms as appropriate.
4	Seeking feedback and advice.
5	Translating and back translating the site-specific instrument.
6	Pilot testing the site-specific instrument.
7	Adapting the data entry templates, data analysis code and reporting templates as appropriate.

Note: Further details on each of these stages are provided in the following pages.

Available support

The WHO Geneva STEPS team is available at all stages of this process for consultation and technical advice. To enable the WHO Geneva STEPS team, to assist with data entry, analyses, and weighting of the data, please ensure that they receive a copy of the Instrument prior to finalization.

Continued on next page

Adapting the STEPS Instrument, Continued

Common questions for adaptation

The table below provides some examples of questions in the STEPS Instrument that most commonly require local adaptation:

Question Code	Standard wording	Guidance for adaptation
C5	What is the highest level of education you have completed?	<ul style="list-style-type: none"> • The education categories (taken from the World Health Survey) are designed to translate national education programmes into an internationally comparable set of categories. • If you use other categories you should document the definitions and how they relate to those in the Instrument.
C6	What is your [<i>insert relevant ethnic group/racial group/cultural subgroup/others</i>] background?	<ul style="list-style-type: none"> • Insert a list of terms that best define differences in health and health related outcomes in your country, e.g. race, religious, ethnicity, etc. • Base ethnic groups on the census definition.
C8	Which of the following best describes your main work status over the past 12 months?	<ul style="list-style-type: none"> • Insert categories appropriate to your setting. • Document the list of the new categories and how they relate to the Instrument.
C11	If you don't know the amount, can you give an estimate of the annual household income if I read some options to you? Is it less than... [Insert Quintile Values]?	Insert 20, 40, 60, 80% of average national income distribution obtained from an authentic source (e.g. National Income and Expenditure surveys, etc).
T1	Do you currently smoke any tobacco products, such as cigarettes, cigars or pipes?	Develop a show card that covers all tobacco products used in your country (see example in Part 5, Section 3).

Note: For further guidance and details about each item in the STEPS Instrument, please see the Question-by-Question Guide in Part 5, Section 2.

Skip patterns and question numbers

If the content of the Instrument has been adapted, you will need to review and update all the skip instructions and question numbers to ensure they are accurate.

Note: Currently the skip instructions reflect the codes, but it may be easier for the interviewers to change these to the finalized question numbers.

Continued on next page

Adapting the STEPS Instrument, Continued

Adapting forms, procedures and show cards

Some forms, procedures and show cards may also require tailoring to ensure local relevance.

The table below shows some common adaptations that may be required.

Item	What to adapt (or create)
Show Cards	Adapt (or create) examples used for: <ul style="list-style-type: none">• list of work status;• list of tobacco products;• standard drink sizes for alcohol consumption;• local fruit and vegetables with standardized servings;• physical activities. See Part 5, Section 3 for examples.
Interview Tracking Form	May require adjustment according to variations in sampling design.

Note: The Interview Tracking Form needs to be used during the interview process. This form is needed to weight the data during data analysis.

Translating STEPS Documents

Introduction

Many sites will require that the WHO STEPS manual and associated documents are available in more than one language. These materials are to be translated into the language(s) used in the sites by a translator and then back-translated into the original language by a different translator to ensure accurate reproduction of meanings.

Documents to translate

The table below lists some of the documents that may need translating and includes the reference to their location in the manual.

Documents	Manual reference
STEPS Instrument	Part 5, Section 1
Question-by-Question Guide	Part 5, Section 2
Show Cards	Part 5, Section 3
Training and Practical Guides	Part 3
Interview Tracking Form, Clinic Registration Form	Part 6, Section 2
Participant Information Form	Part 6, Section 2
Consent forms	Part 6, Section 2

Purpose

The purpose of translation and back-translation is primarily to produce a locally-understandable site-specific instrument and all supporting documents and that the original intent of the questions is maintained.

This will ensure that all interviewers ask the questions in a standardised way and all STEPS documents are clear and understandable to participants.

Language selection

There may be several recognised languages within a country. In this situation:

- interviewing materials may need to be translated into each of these
- trained translators and interviewers will have to be available.

Notes:

- Check if another country/site has already translated the STEPS Instrument into your local language and is willing to share it.
 - Your census office or another government department may help with determining other languages you need to use.
-

Continued on next page

Translating STEPS Documents, Continued

Translation process

Follow the guidelines below to select appropriate translators and ensure accurate and appropriate translation of the site-specific instrument and all other interviewing materials.

- Initial translation of material should be conducted by at least one translator, ideally a linguistic expert who can explain the terms used and suggest alternatives and has experience in health surveys.
 - The instrument must then be back-translated into the original language by another translator to ensure accurate reproduction of meanings.
 - Do not use ‘interpreters of convenience’, such as members of the participant’s family or household, the village headman or any other convenient person present, as it may lead to incorrect data being recorded.
-

Quality standards for translation

Recommended guidelines for translation are listed below.

- Translate the original intent of the questions with the most appropriate equivalent term in the local language.
 - Develop an inventory of local expressions used as well as comparisons of expressions in other languages.
 - Where there are many dialects and/or languages that are not available in written format, carefully plan specific translation protocols.
-

Pilot Testing

Introduction A pilot test of the entire data collection process must be conducted among a limited number of people with a broad range of backgrounds prior to implementing the actual survey. It involves all aspects of the survey including:

- approaching potential participants
- seeking and obtaining informed consent
- making arrangements/appointments for data collection
- site preparation and set-up
- collecting all needed data
- identifying participants who may need follow-up.

Additionally, a pilot test of the entire data entry process must be conducted prior to the actual start of data entry. More information about the data entry pilot test is available in Part 3, Section 5.

When to conduct pilot test Ideally, the pilot test should be conducted as soon as the translated versions of the site-specific instrument and other interview materials are ready.

In practical terms, however, it is recommended that it be conducted after the recruitment and training of data collection staff so trained interviewers can be used during the pilot. This will ensure interviewer consistency and test interviewer skill prior to the main survey.

Test group Identify and approach willing participants to be part of the pilot test. The test group should include the following:

- 10 - 20 people
 - both men and women
 - cover age range used in STEPS
 - more than one ethnic group (if appropriate)
 - people with differing levels of education
 - people from a range of socio-economic groups.
-

Test environment Where possible conduct the pilot test under realistic field conditions.

Timeframe When planning the pilot test, allow sufficient time for adjustments to be made prior to starting data collection.

Continued on next page

Pilot Testing, Continued

Conducting the pilot test Follow the steps below to conduct the pilot test with each participant.

Step	Action
1	Briefly explain the purpose and aim of STEPS chronic disease risk factor surveillance.
2	Briefly explain the purpose of the pilot test.
3	Get each participant to read and sign the necessary consent forms.
4	Using the site-specific instrument, conduct the interviews and record results.

Feedback At the end of each interview, ask the participant the following questions and record their feedback:

- Did any of the questions make you feel uncomfortable?
 - Did you understand all the words?
 - How clear was the intent of the questions?
 - Did you know what was being asked?
 - How could we make it clearer?
 - How else could we improve this survey?
-

Evaluation and refining the Instrument On completion of the pilot test:

- compile all participants' comments into a single report;
 - where necessary, adapt and refine the instrument - taking care not to change intended meanings;
 - send the instrument to WHO Geneva STEPS Team for comments and quality assurance.
-

Section 2: Preparing the Sample

Overview

Introduction This section covers the principles, methods, and tasks needed to prepare, design, and select the sample for your STEPS survey.

Intended audience This section is primarily designed to be used by those fulfilling the following roles:

- statistical adviser
 - STEPS Site Coordinator
 - STEPS Coordinating Committee.
-

Tasks and timeframes The sample is prepared as part of the process of planning and preparing the survey. This process should take between two days to one week, depending on the methods chosen and availability of information needed to draw the sample.

The chart below lists the main tasks and timeframes covered in this section.

Task Name	Duration	Month 1
Define target population	1 day	
Determine sample size	1 day	
Identify sample frame and design	1 week	
Select sample participants	3 days	
Document sample selection	1 day	

In this section This section covers the following topics:

Topic	See Page
Sampling Guidelines	2-2-2
Determining the Sample Size	2-2-3
Identifying the Sampling Frame	2-2-10
Choosing the Sample Design	2-2-12
Selecting the Sample	2-2-20
The Kish Method	2-2-24
Documenting the Sample Design	2-2-26
Preparing Data Collection Forms	2-2-27

Sampling Guidelines

Introduction High quality survey techniques can provide a good picture of risk factors for chronic diseases in a population by using a sample of that population. This is achieved by scientifically selecting the sample from the population. The sample will represent the entire target population if the sample is drawn correctly (91). High standards of sample design and selection are essential to achieve valuable and useful results from STEPS.

Reflecting the scope of your survey in your sample To achieve a sample that reflects the scope of the survey you need to:

- define a target population;
- scientifically select a sample of the population that is representative of the target population;
- plan ahead for reporting of survey results by sex and desired age groups.

Define the target population Each site needs to define the target population for their STEPS Survey. To define your population you need to take into account the purpose and use of the survey data. For example, do you need the survey to be representative of the entire population or a specific region?

It is recommended that the target population for STEPS chronic disease risk factor surveillance be at minimum all adults aged 25 to 64 residing in the survey area. The age range may be expanded to include additional age groups, but it is not recommended to have a smaller age range.

Sample population The sample population is a scientifically selected subset of the target population. Once you have defined the target population you select your sample of participants within the target population.

Estimates for age-sex groups The prevalence of most chronic disease risk factors tends to increase with age and vary by sex. Therefore it is recommended that survey results include estimates for specific age groups for each sex, in addition to the total survey population estimates, in order to provide a more nuanced picture of the prevalence of chronic disease risk factors in your target population.

To ensure that precise estimates for each age-sex group can be calculated from the survey data, the total number of age-sex groups must be taken into consideration when calculating the sample size. Reporting estimates for a greater number of age groups will require a larger sample size. While the recommended size of the age groups is 10 years (i.e. 25-34, 35-44, etc.), 20 year age groups may be used if resources are limited. If resources are extremely limited, estimates may be obtained only for the entire age span of the survey (e.g. 25-64). The next topic includes instructions for how to incorporate the total number of desired estimates into the calculation for sample size.

Determining the Sample Size

Introduction

In order to ensure a sufficient level of precision of the survey results, an adequate sample must be drawn from the target population. To calculate the sample size needed, the following factors must be taken into consideration:

- desired level of confidence of the survey results
- acceptable margin of error of the survey results
- design effect of the sampling methodology
- estimated baseline levels of the behaviours or indicators we want to measure.

Additionally, the sample size must be adjusted for:

- number of age-sex estimates
 - anticipated non-response.
-

Helpful Terminology

The following table provides a brief description of several key statistical terms. It is important to develop a good understanding of this terminology before proceeding to calculate the sample size.

Term	Description
Sample Mean / Prevalence	The estimated mean or prevalence of a given population parameter (e.g. mean number of days fruit was consumed in a given week) that is calculated from the survey data.
Population Mean / Prevalence	The true mean or prevalence of a given parameter for the entire target population. The sample mean is an estimate of the population mean.
Confidence Intervals	A range of values around the sample mean or prevalence in which the population mean or prevalence is likely to fall. For example, a 95% confidence interval indicates that for 95 out of 100 surveys, the population mean would fall into this range of values around the sample mean.

Continued on next page

Determining the Sample Size, Continued

Variables used for calculating sample size

The table below provides a description of the variables used in calculating the sample size as well as the recommended values for each variable.

Variable	Description	Recommended Value
Level of Confidence	<ul style="list-style-type: none"> • Probability value that is associated with a given confidence interval. • Describes the level of uncertainty in the sample mean or prevalence as an estimate of the population mean or prevalence. • The higher the level of confidence, the larger the sample size needed. 	<ul style="list-style-type: none"> • 1.96 • Note: 1.96 is the probability value associated with a 95% confidence interval.
Margin of Error	<ul style="list-style-type: none"> • The expected half-width of the confidence interval. • The smaller the margin of error, the larger the sample size needed. 	<ul style="list-style-type: none"> • 0.05 • Note: If the estimated baseline levels of the behaviours or indicators you wish to measure is very low (e.g. <0.10), then the Margin of Error should be decreased to 0.02 or smaller.
Design Effect (Deff)	<ul style="list-style-type: none"> • Describes the loss of sampling efficiency due to using a complex sample design. • The design effect for a simple random sample is 1.00. Sample designs more complex than a simple random sample require a larger sample to achieve the same level of precision in survey results as a simple random sample. Thus the design effect increases as the sample design becomes more complex. 	<ul style="list-style-type: none"> • 1.50 • Note: The value 1.50 is recommended for most STEPS surveys with complex sample designs. If design effect information is available from previous national surveys of a similar design to the proposed STEPS survey, it is recommended to use the previous estimates for design effect.
Estimated baseline levels of the behaviours or indicators we want to measure	<ul style="list-style-type: none"> • The estimated prevalence of the risk factors within the target population. • Values closest to 50% are the most conservative, requiring the largest sample size. 	<ul style="list-style-type: none"> • 0.50, if no previous data are available on the target population. • The value closest to 0.50, if previous data is available on the target population.

Equation for calculating sample size

The equation for calculating sample size is as follows:

$$n = Z^2 \frac{P(1-P)}{e^2}$$

where:

- Z = level of confidence
- P = baseline level of the indicators
- e = margin of error

Determining the Sample Size, Continued

Example calculation

Using the above recommendations for each variable, the **initial** calculation for sample size would be:

$$n = 1.96^2 \frac{0.5(1-0.5)}{0.05^2} = 384$$

However, this number **must** be adjusted to account for the design effect of the sample design, the number of age-sex estimates to be reported, and the anticipated non-response.

Adjusting for design effect

To adjust for the design effect of the sample design simply **multiply** the sample size by the design effect. For more information on choosing the sample design for your survey, see page 2-2-12.

Adjusting for number of age-sex estimates

As discussed previously, it is recommended that survey results be reported separately for specific age groups for each sex. In order to have an adequate level of precision for each age-sex estimate, the sample size must be **multiplied** by the number of age-sex groups for which estimates will be reported.

The number of age-sex estimates will vary according to the target age range of the survey and the resources available for the survey. For surveys covering the age range of 25-64, the recommended number of age-sex estimates is **8**, or 4 10-year age groups per sex. However, if resources are limited the number of age-sex estimates can be reduced to 4 (e.g. 20-year age groups for each sex) or 2 (e.g. 40-year age groups for each sex).

If the age range of your survey extends beyond the recommended 25-64, the total number of age-sex estimates may need to be adjusted accordingly. For example, if the age range of 15-24 were also to be included in the survey and 10-year age-sex estimates are desired, the total number of age-sex estimates would be 10.

Adjusting for anticipated non-response

To adjust for anticipated non-response **divide** by the anticipated **response rate**.

A non-response rate of 20% is the recommended rate to anticipate. This is a conservative estimate based on response rates of previous STEPS surveys. If response rates have been consistently higher at your site for similar household surveys, a less conservative (i.e. smaller) non-response rate may be used, such as 10%.

Example: For an anticipated non-response rate of 20%, divide the sample size by 0.80.

Continued on next page

Determining the Sample Size, Continued

Summary of sample size calculation

The table below provides a summary of the above steps to calculate sample size.

Step	Description
1	Determine the value of all variables needed to calculate sample size.
2	Use the level of confidence, margin of error, and baseline level of the indicators in the above equation to get an initial estimate for n (sample size).
3	Multiply n by the design effect and by the number of age-sex estimates.
4	Divide the result from step 3 by the anticipated response rate to attain the final sample size.

Sample Size Calculation Example 1

In this example, the recommended values for all parameters of the sample size equation will be used. Thus, the initial calculation proceeds as follows:

$$n = 1.96^2 * \frac{0.5(1-0.5)}{0.05^2} = 384$$

This initial n is then multiplied by the design effect of 1.5 and the 8 age-sex estimates desired for the survey results:

$$n = 384 * 1.5 * 8 = 4,608$$

Finally, n is divided by 0.80 to adjust for the anticipated 20% non-response rate:

$$n = 4,608 \div 0.80 = 5,760$$

5,760 is the final sample size.

Sample Size Calculation Example 2

In this example, the recommended values for all parameters of the sample size equation will be used and the initial calculation proceeds just as in the previous example:

$$n = 1.96^2 * \frac{0.5(1-0.5)}{0.05^2} = 384$$

However, in this example the estimates will only be reported for 20-year age groups for each sex as the sample size required for 10-year age groups is too large for the resources available. Thus, the initial n is then multiplied by the design effect of 1.5 and 4 age-sex estimates desired for the survey results:

$$n = 384 * 1.5 * 4 = 2,304$$

Finally, n is divided by 0.80 to adjust for the anticipated 20% non-response rate:

$$n = 2,304 \div 0.80 = 2,880$$

2,880 is the final sample size.

Continued on next page

Determining the Sample Size, Continued

Sampling very small populations

When the target population is very small (appx. <50,000 people) the sample size can be reduced using a Finite Population Correction (FPC). The steps below describe how to check if the FPC is appropriate for your site and how to apply it to reduce your sample size.

Step	Description										
1	Complete only steps 1 and 2 in the preceding table to obtain the n for each estimate.										
2	Calculate the target population size for each estimate using available census data or a similar reliable data source. Example: If 8 10-year age-sex groups will be the estimates, the number of individuals in each age-sex group (e.g. number of males aged 25-34) must be calculated.										
3	The FPC should only be applied when the sample to be drawn represents more than 10% of the target population. Thus for each estimate the n calculated in Step 1 must be divided by the target population size for that estimate to check to see if the FPC can be applied. Example: n has been calculated as 384. Eight 10-year age-sex estimates are desired. The table below shows the data collected for the first four estimates. <table border="1" data-bbox="544 1133 1222 1328"> <thead> <tr> <th>Desired Estimates</th> <th>Target Population Size</th> </tr> </thead> <tbody> <tr> <td>Males, 25-34</td> <td>2548</td> </tr> <tr> <td>Females, 25-34</td> <td>2641</td> </tr> <tr> <td>Males, 35-44</td> <td>3465</td> </tr> <tr> <td>Females, 35-44</td> <td>3356</td> </tr> </tbody> </table> <p>Divide n by the target population for each estimate: $384/2548 = 0.15$ $384/2641 = 0.15$ $384/3465 = 0.11$ $384/3356 = 0.11$</p>	Desired Estimates	Target Population Size	Males, 25-34	2548	Females, 25-34	2641	Males, 35-44	3465	Females, 35-44	3356
Desired Estimates	Target Population Size										
Males, 25-34	2548										
Females, 25-34	2641										
Males, 35-44	3465										
Females, 35-44	3356										
4	If most or all of the quotients from step 3 are 0.10 or higher, then the FPC can be applied (continue to next step). Otherwise, return to step 3 in the preceding table and continue to calculate the total sample size using the n already calculated.										
5	Apply the FPC to the n for each estimate using the following equation: $\text{new } n = \frac{n}{1 + \frac{n}{\text{population}}}$ <p>where "population" refers to the target population for a given estimate, not the entire target population.</p>										

Continued on next page

Determining the Sample Size, Continued

Sampling very small populations (cont.)

Step	Description
6	Sum all the "new <i>n</i> 's" together and multiply the sum by the design effect.
7	Divide the result from step 6 by the anticipated response rate to attain the final sample size.

Further modifications to sample size

There are a variety of situations which may require an adjustment to the sample size resulting from the calculations above. The table below describes some of these situations with directions on how to adjust the sample size. If you do not see your situation listed here or if any other additional assistance is required, please contact the STEPS team.

If ...	Then ...						
Data for specific subgroups are required (e.g. ethnic groups, urban vs. rural dwellers).	There are two ways to proceed depending on the information desired:						
	<table border="1"> <thead> <tr> <th>If ...</th> <th>Then ...</th> </tr> </thead> <tbody> <tr> <td>Data will only be reported for all individuals in each subgroup.</td> <td>Set the number of estimates to the larger of: <ul style="list-style-type: none"> the number of age-sex estimates desired the number of new subgroups. </td> </tr> <tr> <td>Data will be reported for each age-sex group within each subgroup.</td> <td>Multiply the number of age-sex groups by the total number of new subgroups (e.g. total number of ethnic groups) to determine the total number of estimates.</td> </tr> </tbody> </table>	If ...	Then ...	Data will only be reported for all individuals in each subgroup.	Set the number of estimates to the larger of: <ul style="list-style-type: none"> the number of age-sex estimates desired the number of new subgroups. 	Data will be reported for each age-sex group within each subgroup.	Multiply the number of age-sex groups by the total number of new subgroups (e.g. total number of ethnic groups) to determine the total number of estimates.
	If ...	Then ...					
Data will only be reported for all individuals in each subgroup.	Set the number of estimates to the larger of: <ul style="list-style-type: none"> the number of age-sex estimates desired the number of new subgroups. 						
Data will be reported for each age-sex group within each subgroup.	Multiply the number of age-sex groups by the total number of new subgroups (e.g. total number of ethnic groups) to determine the total number of estimates.						
Note: It is important to take these subgroups into mind when allocating the sample to ensure a sufficient number of participants can be drawn from each subgroup (see next topic).							
Oversampling is desired for very small sub-populations.	Increase the overall <i>n</i> by increasing the <i>n</i> for the specific estimate(s) by 10%.						
Oversampling is desired for specific sub-populations with higher than average non-response.	Increase the overall <i>n</i> by increasing the <i>n</i> for the specific estimate(s) by 10 to 20%.						

Continued on next page

Determining the Sample Size, Continued

Further modifications to sample size (cont.)

If ...	Then ...
Oversampling of the 55-64 age group is desired because obtaining sufficient numbers of respondents from this age group is expected to be difficult due to high non-response and/or small size of this sub-population.	Increase the overall n by increasing the specific estimates for males and females in this age group by 10 to 20%. See the discussion of the Kish Method, beginning on p. 2-2-24, for specific instructions for oversampling 55-64 year olds within households.

Note: If oversampling is desired, adjustments usually must also be made when allocating the sample (see next topic). Often in addition to increasing the sample size, the sample allocation must take into consideration the location of hard-to-reach groups and allocate a greater proportion of the sample to these areas.

Sample Size Calculator

There is an Excel workbook, [sample_size_calculator.xls](#), that can assist you in the calculations needed to determine the sample size for your survey. It is available on the STEPS CD and STEPS website. The calculator allows you to adjust all variables discussed here and also provides assistance in determining whether the Finite Population Correction (FPC) is applicable to your survey and, if so, how to correctly apply the FPC.

Smaller sample sizes

If the sample size calculations result in a sample size too large for the resources available, consider reducing the number of age-sex estimates desired for your results. Reducing the age-sex estimates from 10-year age groups to 20-year age groups can significantly reduce the sample size required for your survey.

An additional means to save costs is to only conduct Step 3 on a subsample of the participants for Steps 1 and 2. However, this will reduce the precision of the population estimates and smaller age ranges should be used in reporting Step 3 results. If subsampling for Step 3 is done, a minimum of 20% of the total sample size should be targeted for Step 3.

Identifying the Sampling Frame

Introduction

A sampling frame is a list of units or elements that defines the target population. It is from this list that the sample is drawn. A sampling frame is essential for any survey.

Finding available sampling frames

To identify available sampling frames and determine which is best for your site, search for updated lists, databases, registers or other sources that give good coverage of the population you wish to survey. For example, look for population registers or census lists.

Various government departments and national bodies should be consulted to establish what frames exist in your country and, if suitable, whether they may be accessed for STEPS.

Enumeration areas (EAs)

Most often the sampling frame will use enumeration areas (EAs) which are small- to medium-sized geographic areas that have been defined in a previous census. Most countries have this information and it is usually preferable to incorporate this into the sampling frame.

Factors to consider

A sampling frame, or a collection of them, should cover all of the population in the surveyed site. Good coverage means that every eligible person in the population has a chance of being included in the survey sample.

Representativeness for all sub-populations should be considered when deciding which frame(s) to use. You need to watch out for the possibility that particular age, gender or ethnic groups or geographical areas are more or less likely to be included in the sampling frame. Bias will occur if there is poorer coverage for some groups.

Multiple Sampling Frames

Due to logistical and financial limitations, most national surveys employ multi-stage sampling, which is discussed in detail in the following topic. A multi-stage sample design will require a sampling frame for each stage of sampling.

Continued on next page

Identifying the Sampling Frame, Continued

Features of a good sampling frame

Some features of a good sampling frame are:

- it does not contain duplicates, or if present they can easily be identified and removed;
- it does not contain blanks, such as empty houses or a deceased individual;
- it contains information enabling all units to be distinguished from all others and to be easily located (e.g. a complete street address);
- at minimum, it contains information about the number of households or total number of individuals;
- it could be made accessible to the STEPS team within a reasonable timeframe and at no large expense.

Note: Sampling frames must be assessed for all the above features, but particularly for **completeness** and **potential bias**.

Choosing the Sample Design

Introduction The selection of the sample design is highly dependent on a variety of factors, most importantly the size of the population, the geography of the area to be covered, and the resources available for the survey. All factors must be kept in mind in selecting the sample design for the survey.

Stratification Stratification is the process of dividing the sampling frame into mutually exclusive subgroups or strata. The sample is then drawn either proportionately or disproportionately from **all** strata. How the target population is stratified depends on the information that is available for the sampling frame and the information that is desired from the survey results.

Strata are often based on the physical location of the sampling units. Some examples of these types of strata are:

- enumeration areas (EAs) or other well-defined geographic regions
- urban vs. rural areas.

Less often, strata are based on the characteristics of the individuals in the sampling frame. This is less common in large national surveys due to a lack of precise data on all individuals in the target population and the difficulties of developing sampling frames for each strata. Some examples of these types of strata are:

- ethnicity
- socioeconomic status
- gender.

Stratification is not required but is recommended for the following reasons:

- increased precision of survey estimates
- guaranteed coverage of all strata
- administrative convenience.

Stratification can be applied in conjunction with other sampling strategies. This section discusses simple random sampling and multi-stage cluster sampling, both of which can be used along with stratification, as described later in this topic.

Stratification and sample allocation If the decision has been made to stratify the population, it must then be decided whether to sample proportionately from all strata or to sample a larger proportion of individuals from some strata and a smaller proportion of individuals from other strata (disproportional allocation).

Continued on next page

Choosing the Sample Design, Continued

Stratification and sample allocation (cont.)

Proportional allocation means sampling the same proportion of individuals from each strata so that the resulting sample is distributed across the strata similarly to the underlying target population. This type of sample allocation is the appropriate method for surveys which will only be reporting data for all strata combined.

Disproportional allocation means sampling some strata at a higher rate than other strata. Often this is implemented by drawing an equal sized sample from each strata. This type of sample allocation is appropriate when survey results are desired for each individual strata. In this situation, a larger sample size is usually required to ensure adequate precision in the strata-specific estimates. The primary drawback to this method is a loss of sampling efficiency for the estimates for all strata combined.

Note: In some cases where very small strata exist, proportional allocation may be done but oversampling may be required for the very small strata.

Proportional Allocation Example

Because proportional allocation is more likely to be used for a STEPS survey, an example is provided here.

In this example, the sample size has been calculated to be 3,000. The target population has been divided into the 4 government districts of the country. These districts will serve as strata. The target population within each strata has been listed in the table below along with the proportion each comprises of the total target population.

Strata	Target Pop.	Proportion of Pop.
District 1	25,955	0.24
District 2	30,568	0.28
District 3	32,578	0.30
District 4	19,054	0.18
Total	108,155	1.00

$= 25,955 \div 108,155$

To compute the number of individuals from the total sample to be drawn from each strata, multiply the total sample size by the proportion for each strata.

Strata	Target Pop.	Proportion of Pop.	Sample
District 1	25,955	0.24	720
District 2	30,568	0.28	840
District 3	32,578	0.30	900
District 4	19,054	0.18	540
Total	108,155	1.00	3,000

$= 0.24 \times 3,000$

Continued on next page

Choosing the Sample Design, Continued

Simple random sampling

In a small number of settings simple random sampling may be feasible. For household surveys, the following characteristics generally should be met:

- small target population;
- small survey area, the entirety of which can be covered by the resources available;
- detailed sampling frame is available, listing, at minimum, all households in the survey area, or, at best, all eligible individuals in the survey area.

Simple random sampling can be combined with stratification. In stratified random sampling, the population is first stratified and then a random sample is drawn from each strata.

Note: If simple or stratified random sampling is deemed to be feasible at your site, a smaller sample size can be used. In the calculation for sample size a design effect of 1 should be used.

Multi-stage cluster sampling

Multi-stage cluster sampling is one of the most common sample designs for national surveys and it is the recommended method for most STEPS surveys.

"Multi-stage" indicates that sampling is done in several steps. First larger sampling units are selected then smaller sampling units are selected within the selected larger units. "Cluster" refers to the fact that the sampling units are subdivided into mutually-exclusive clusters and, unlike stratification, only a **sample** of these clusters is selected for the survey.

Why use multi-stage cluster sampling?

The table below highlights two primary reasons for using multi-stage cluster sampling. These are very common problems in national surveys that can be overcome with the use of multi-stage cluster sampling.

Problem	Solution
Detailed information does not exist for all households or individuals in the sample population and it is not feasible to create a detailed sampling frame for the entire survey area.	Multi-stage cluster sampling allows for the selection of larger sampling units (e.g. villages) that require less detailed information about the target population. It is only at the final stage of sampling (most often the selection of households) that detailed information needs to be available. However, because only a selection of clusters will be chosen at each stage of sampling, the detailed sampling frames are only needed for a subset of the entire target population.

Continued on next page

Choosing the Sample Design, Continued

Why use multi-stage cluster sampling? (cont.)

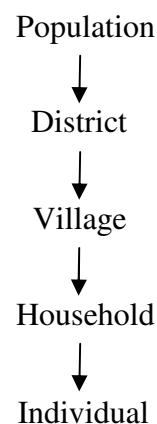
Problem	Solution
The survey area is too large and/or travel costs are too high to draw a sample from the entire country or all regions of interest.	<p>Because the sample is only drawn from selected clusters, multi-stage cluster sampling allows for a reduced area to be surveyed while maintaining a sample that is nationally (or subnationally) representative.</p> <p>Note: Using multi-stage cluster sampling does not <i>guarantee</i> a representative sample. If done incorrectly, it will not result in a representative sample. The design of the clusters and the selection of clusters at every stage must be done carefully and consistently and must be documented in detail.</p>

Preparing a Multi-stage Cluster Sample

In order to implement multi-stage cluster sampling, the population must be divided into clusters, each of which contain either a number of smaller clusters or, at the final stage, households or individuals.

The flowchart to the right is one example of the multiple sampling stages that could be defined for a site.

Most often the first stage uses enumeration areas (EAs) from census information. The intermediary stages, if any, may be comprised of existing geopolitical units (e.g. villages) or artificially-created units (e.g. a specified collection of city blocks).



Important: The number of sampling units at the initial stage must be fairly numerous (i.e. >100) so at least 50-100 of them can be selected. Selecting a smaller number of sampling units at the initial stage of sampling results in more clustered data and a loss of precision in survey estimates.

A sampling frame will need to be constructed for all clusters in the first stage of sampling. At minimum these sampling frames must contain the total number of households or total number of target individuals in the cluster.

Sampling frames will only be needed for **selected** clusters at all subsequent stages of sampling, with detailed information (i.e. lists of households or eligible individuals) only needed for the sampling frames for the last stage of sampling.

Choosing the Sample Design, Continued

Multi-stage Cluster Sampling Terminology

The table below describes some key terminology for multi-stage cluster sampling.

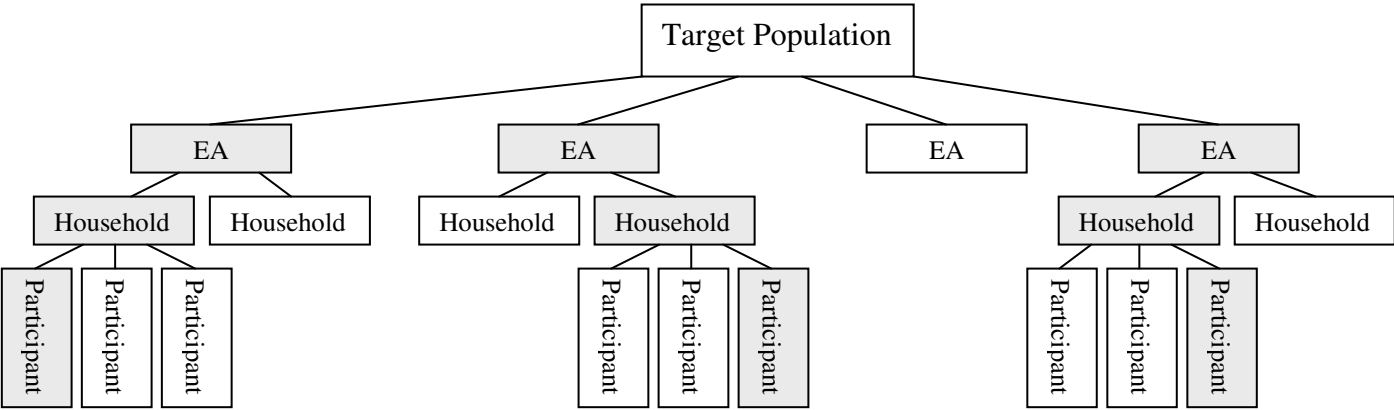
Term	Definition
Primary Sampling Unit (PSU)	These are the clusters that are selected first. Most often the PSUs are enumeration areas (EAs) from a recent census.
Secondary Sampling Unit (SSU)	The clusters that are selected second, separately within each selected PSU.
Tertiary Sampling Unit (TSU)	The clusters that are selected third, separately within each selected SSU.

The list of terms could be extended to describe more levels of sampling as needed.

Example 1

In the following example, there are three stages of sampling. EAs are serving as the PSUs. For each selected PSU, a sampling frame was created comprised of a list of households in the EA. Households were then selected within each PSU and then one participant was selected within each household.

Shaded boxes indicate that the cluster or participant was selected.



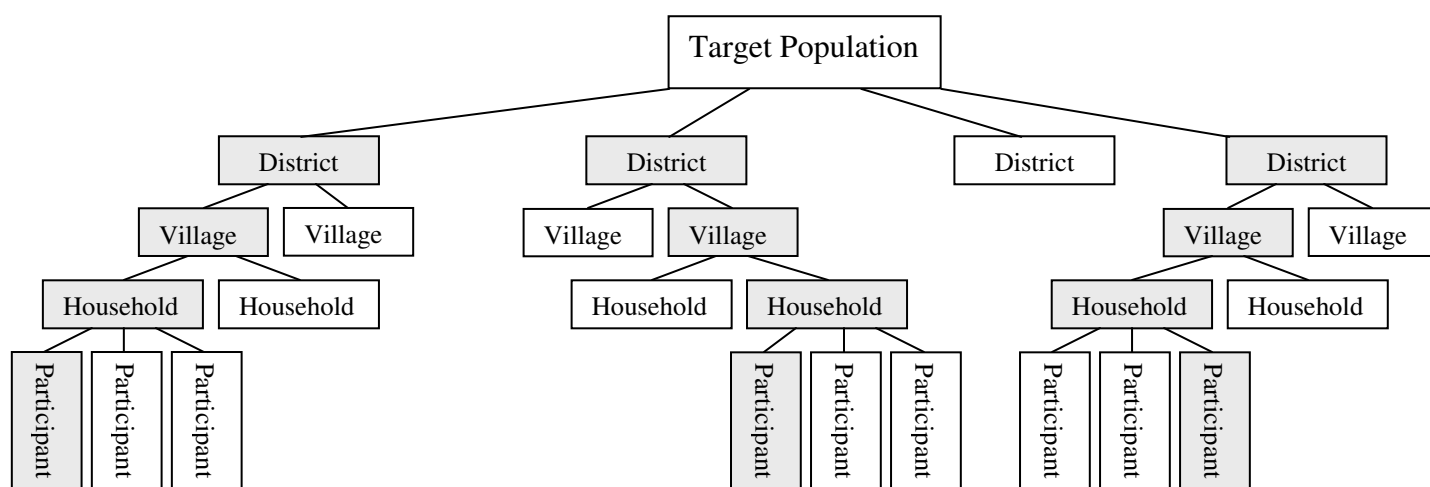
Continued on next page

Choosing the Sample Design, Continued

Example 2

In this example, there are four stages of sampling. Districts are serving as the PSUs. For each selected PSU, a sampling frame was created comprised of a list of all villages (the SSUs) with the target population of each village. For each selected village, a sampling frame was also created, comprised of a list of all households in the village. If a detailed list of all eligible individuals were available for any selected village, this list could be used in place of the household list and selection could proceed directly from the village level to the participant level.

Shaded boxes indicate that the cluster or participant was selected.



Qualities of a Good Multi-stage Cluster Design

One very important check to perform on the multi-stage cluster design is that every individual in the target population is included in only one sampling unit per stage. This means that the clusters at each level of sampling must cover the entire target population and be mutually exclusive (non-overlapping).

Additionally, it is important to check the characteristics of the PSUs. The first two items in the table below can be used to check the SSUs, TSUs, etc. as well, but given the nature of multi-stage cluster designs, these checks are most critical for the PSUs.

If ...	Then ...
PSUs exist that are very small.	Combine these PSUs with a neighboring PSU before selecting the sample.
PSUs exist that are very large.	Split these PSUs into two or more smaller PSUs that are more similar in size to other PSUs.
Total number of PSUs is small (i.e. <100).	Begin sampling at the SSU level (the SSUs would then become PSUs) or subdivide the existing PSUs to ensure that at least 50-100 PSUs can be selected.

Continued on next page

Choosing the Sample Design, Continued

Sample Allocation and Multi-stage Cluster Design

Once the sampling units to be used for PSUs, SSUs, etc. have been determined, the allocation of the sample must be decided. That is, the total number of PSUs to be selected, the total number of SSUs to be selected per PSU, etc. must be determined.

The table below describes the steps to take to determine how to allocate the sample.

Step	Description
1	Calculate the total sample size.
2	Assess the resources available and determine the total number of PSUs to be sampled, keeping in mind that at least 50 to 100 PSUs should be selected.
3	Divide the total sample size by the number of PSUs to be sampled to determine the number of individuals to be sampled per PSU.
4	Continue subdividing the sample size at each stage of sampling according to the number of sampling units to be selected at each stage.

Note: As stated previously, stratification can be combined with a multi-stage cluster design. The total number of PSUs would be allocated proportionately or disproportionately (depending on the requirements of the survey results) across all strata and sample allocation would continue within each strata following the steps above.

Example

For this example, assume that the total sample size has been calculated to be 3,200 individuals. It has also been decided that regions will serve as PSUs, villages will serve as SSUs, and then households will be selected in each village. Resources will allow for 80 PSUs to be selected, meaning that 40 ($= 3200/80$) individuals will be selected per PSU.

There is some flexibility in how the 40 individuals per PSU are allocated. At this point it would be worthwhile to consider a few scenarios and select the one that is feasible yet provides a good distribution of individuals across the PSU (i.e. not too many or too few of the 40 individuals drawn from a given village). Two scenarios are presented below:

Continued on next page

Choosing the Sample Design, Continued

Example (cont.)

Scenario	Description
1	10 individuals will be selected per village, meaning that 4 villages (= 40/10) must be selected per PSU. <u>Sample allocation:</u> 80 regions x 4 villages/region x 10 individuals/village = 3200
2	5 individuals will be selected per village, meaning that 8 villages (= 40/5) must be selected per PSU. <u>Sample allocation:</u> 80 regions x 8 villages/region x 5 individuals/village = 3200

In terms of resources, the key difference between the above scenarios is the number of villages that would need to be visited within each PSU. This number will likely be a deciding factor in the allocation of the sample, keeping in mind that having a high number of individuals selected from only a few villages would result in greater clustering of survey data and a potential loss of precision in survey estimates.

Example with stratification

For this example, assume again that the total sample size has been calculated to be 3,200 individuals and that regions will serve as PSUs, villages will serve as SSUs, and then households will be selected in each village. Just as in the previous example, resources will allow for 80 PSUs to be selected. However, in this example, the survey designers wish to ensure that the sample is drawn proportionately across the 4 islands that comprise the country.

The table below shows the proportion of the total underlying population that each island represents. The right-most column shows how the number of PSUs would be proportionately allocated across these 4 islands or strata.

Island	Proportion of Total Pop.	PSUs
A	0.50	40
B	0.175	14
C	0.125	10
D	0.20	16
Total	1.00	80

Thus, 40 regions (PSUs) will be picked out of all regions on island A, 14 regions will be picked out of all regions on island B, and so on. Once the PSUs are selected per island, sample allocation continues just as in the preceding example, with the same number of villages being selected in each PSU, regardless of the island on which the PSU is located.

Selecting the Sample

Introduction Once the sample design is selected and the sampling frame has been prepared, you are ready to proceed with sample selection. This section provides instructions for the various stages of sampling.

Available tools There is an Excel workbook entitled **STEPSsampling.xls** that includes spreadsheets for every stage of the sample selection. STEPSsampling.xls will:

- provide probability proportional to size (PPS) sampling (see description below) for primary and secondary sampling units as needed;
- randomly select households or individuals;
- provide information for weighting the data.

The spreadsheet is available on the STEPS website (www.who.int/chp/steps) and on the CD-Rom.

Probability proportional to size (PPS) sampling Probability proportional to size (PPS) sampling is a method for selecting a sampling unit in which the probability of selection for a given sampling unit is proportional to its size (most often the number of individuals or households within the sampling unit).

PPS sampling is appropriate for use when sampling units are of markedly different size. In these situations, were random sampling to be used to select sampling units, those individuals in the larger sampling units would have a much smaller chance of selection than those individuals in the smaller sampling units. PPS sampling corrects this problem, therefore reducing bias in survey estimates.

Instructions for PPS sampling The table below outlines the steps required to perform PPS sampling on a list of sampling units. Before beginning, a list of sampling units and their corresponding sizes (in number of households or in population) must be compiled. It is recommended that this list be organized geographically, meaning that sampling units located near each other are also near each other on the list. Additionally, the number of sampling units (clusters) to be selected must be decided.

The **STEPSsampling.xls** tool will automatically perform Steps 3 through 8 in the table below. The instructions worksheet inside the file explains how to perform PPS sampling using either the PSU or SSU worksheet in the file.

Continued on next page

Selecting the Sample, Continued

Instructions for PPS sampling (cont.)

Step	Action
1	Create a list of all sampling units with their size (either number of households or population). If possible, order this list geographically, placing sampling units that are physically adjacent near each other on the list.
2	Determine the number of sampling units to be selected from the list.
3	Create a new column containing the cumulative size of the sampling units. The final total should match the total population across all sampling units.
4	Divide the total cumulative population size (N) by the number of sampling units to be selected (n) to obtain the sampling interval (k). $k = N/n$
5	Choose a random number (r) that is between 1 and the sampling interval (k). $1 < r < k$
6	Start at the top of the list and select the first sampling unit whose cumulative population size includes the random number (r).
7	To select the second cluster, first add the sampling interval to the random number (r). Then begin counting from the previous cluster selected until the cumulative population size includes this sum (r+k).
8	Select the remaining clusters by adding the sampling interval, multiplied by 2, then 3 and so on, to the random number. Always start counting from the previous cluster selected not the start of the list. $r+(k \times 2)$ $r+(k \times 3)$ etc
9	Continue until the end of the list is reached. Do not stop as soon as n units have been selected. To avoid bias, all units selected must be used in the survey even if the number is slightly greater than n.

Using PPS sampling with a multi-stage cluster design

PPS sampling can be applied at all stages of a multi-stage cluster design except for the final stage in which households or individuals are selected.

The **STEPSsampling.xls** tool provides worksheets for selecting your PSUs and SSUs using PPS sampling. The worksheet entitled PSU allows for the selection of up to 100 PSUs from an entered list of all PSUs. The worksheet entitled SSU allows for the selection of the SSUs within each selected PSU. Therefore, the SSU worksheet must be duplicated, one for each PSU that was selected, so that an independent selection of SSUs can be performed for each PSU.

Continued on next page

Selecting the Sample, Continued

Selection of households and/or individuals

The final stage of sampling, the selection of households and/or individuals, will depend on the type of information available. The table below describes the possible scenarios for the final stage of sampling and the sample selection process for each.

If ...	Then ...
<p>A list of eligible individuals is available for the selected sampling unit (e.g. village).</p>	<p>First check that the list of eligible individuals meets the following requirements:</p> <ul style="list-style-type: none"> • the list is up to date, for example, people who have moved away or who have died are not included in the list; • the list contains specific information allowing for each selected individual to be located by the interviewers. <p>If both conditions are met, the selection of individuals can be done randomly from the list.</p>
<p>No or limited information is available about the individuals in the selected sampling unit but a list of households exists for the sampling unit.</p>	<p>First check that the list of households meets the following requirements:</p> <ul style="list-style-type: none"> • the list is up to date and each household listed represents a single dwelling; • the list contains specific information allowing for each selected household to be located by the interviewers. <p>If both conditions are met, the selection of households can be done randomly from the list. The Kish Method, covered in the next topic, can then be used to randomly select participants from selected households.</p> <p>If there is a concern that the list may be out of date, it is recommended that the survey team first perform a quick survey of the sampling unit to update the list, noting abandoned/destroyed dwellings, new dwellings, or expanded dwellings (single family into multi-family).</p>

Continued on next page

Selecting the Sample, Continued

Selection of households and/or individuals (cont.)

If ...	Then ...
The number of households is known for the sampling unit but there is no information about their location.	In this situation the sampling unit should be mapped to determine the location of the households. Please contact the STEPS team for more guidance on this method or other alternatives.

In the **STEPSsampling.xls** tool, the "RandHhold" worksheet can be used to randomly select the desired number of participants from a list of eligible individuals or the desired number of households from a list of households.

It is possible that some sampling units have more detailed information available than others. In this case, the above scenarios can be used on a case-by-case basis, meaning in some sampling units with more detailed information individuals may be selected directly while in other sampling units with less detailed information households may need to be selected first.

Note: In all STEPS survey designs, sampling is non-replacement, meaning that once a unit or person is selected they are not replaced with another person/unit. If you replace non-respondents or persons who are not at home for the interview you will be performing a convenience sample and your results will only represent the people sampled and not your target population.

The Kish Method

Introduction The Kish Method is a technique that allows for the random selection of one individual from a household.

The Kish Method can be used for selection within households regardless of the sampling method used to select the households (92).

Materials To use the Kish Method you will need the **Kish Household Coversheet**, which is in Part 6, Section 2 of this manual.

Process The table below provides detailed directions on how to implement the Kish Method in each household. An abbreviated version of these directions is also located at the top of the Kish Household Coversheet.

Step	Description																				
1	Ask for the age and sex of all adults aged 25-64 residing in the household. List these in the empty table on the Coversheet.																				
2	<p>Assign a rank to each adult in the table. The ranks should be consecutive and begin with 1. Assign the ranks according to the following rules:</p> <ul style="list-style-type: none">• first assign ranks to males in order of decreasing age (oldest to youngest);• next assign ranks to females in order of decreasing age. <p>An example is provided here:</p> <table border="1"><thead><tr><th>Sex</th><th>Age</th><th></th><th>Rank</th></tr></thead><tbody><tr><td>F</td><td>45</td><td></td><td>2</td></tr><tr><td>M</td><td>45</td><td></td><td>1</td></tr><tr><td>F</td><td>29</td><td></td><td>4</td></tr><tr><td>F</td><td>32</td><td></td><td>3</td></tr></tbody></table>	Sex	Age		Rank	F	45		2	M	45		1	F	29		4	F	32		3
Sex	Age		Rank																		
F	45		2																		
M	45		1																		
F	29		4																		
F	32		3																		
3	In the Kish Selection Table (at the bottom of the Coversheet), find the column whose heading matches the last digit of the Household ID. In this column find the row whose header matches the total number of eligible adults in the household. The number in the box where this row and column intersect gives the rank of the adult to be interviewed.																				

Continued on next page

The Kish Method, Continued

Preparing materials

The Kish Household Coversheets must be prepared prior to beginning data collection. Directions on completing the Coversheets, including how to assign Household ID numbers, can be found in the last topic of this Section, "Preparing Data Collection Forms", on page 2-2-27.

Oversampling for 55-64 year olds

Depending on your site's population structure it may be difficult to obtain enough respondents from the 55-64 year old age group to get precise estimates for this age group. One possible solution to this problem is to oversample this age group at the household level.

At each household with adults aged 55-64, two adults will be selected and two Kish Household Coversheets will be needed. One Coversheet will be used to select one adult from the non-oversampled group and the other will be used to select one adult from the oversampled group. Thus on one Coversheet all adults aged 25-44 will be listed and on the other all adults 55-64 will be listed.

Note: When oversampling be sure to adjust the household size to reflect this. If there are five people in a household and one is 55-64, then the household size for sampling from the 25-44 group is only four, while the individual in the 55-64 age group would have a household size of only one.

Documenting the Sample Design

Introduction Once the sample design and methodology have been chosen, all aspects of the sample need to be clearly documented.

Purpose The purpose of documenting the sample design is primarily for the data analyst to understand how the sample was drawn in order to appropriately adjust the results to the target population. Additionally, an abbreviated version of the documentation should always accompany any presentation of the survey data to explain how the data were collected.

Recordkeeping during data collection Sufficient records must be kept **during data collection** to ensure that the data analyst can do all possible adjustments to make the results representative of the target population. Most importantly, the data analyst must know:

- the probability of selection of each sampling unit at every stage of sampling (i.e. probability of selection for each PSU, SSU, household, individual);
- the age and sex of any non-responders.

Thus, it is critical to keep a record of the following:

- all sampling frames used at each stage of sampling
 - sample selection method used at each stage of sampling
 - stratification design, if stratification is used
 - for each respondent, the PSU, SSU, etc. from which he/she was selected.
-

Future surveys Documenting the sample design and methodology is also important for future surveys when changes in risk factors over time are being examined, since methods chosen in future surveys may differ from this one and thus affect comparability.

Archiving documents It is important that all relevant sampling materials be archived. This includes the forms discussed in the next topic of this Section, "Preparing Data Collection Forms", as well as all information used to design and draw the sample.

If the sample is drawn by another government entity (e.g. the Statistics Bureau), be sure to obtain from them all materials and information that were used to draw the sample.

Preparing Data Collection Forms

Introduction

Once the sample has been drawn the Interview Tracking Forms and the Kish Household Coversheets as well as the STEPS Instruments and, if applicable, the Clinic Appointment Cards should be prepared for the data collection team. It is recommended that the data collection supervisor and the statistical adviser collaborate on this task to ensure the forms are correctly filled out and properly organized for data collection.

Assigning Unique Identifiers

Before preparing the data collection forms, ID Numbers must be assigned to all interviewers and to all selected clusters from which households and/or individuals will be selected. Additionally, all households to be selected and all participants to be selected should each be assigned a unique ID. The table below provides further instruction for assigning these ID Numbers:

Variable	Description	Value Range
Interviewer ID	Every interviewer should be assigned a unique ID number.	1-999
Cluster ID	A unique number should be assigned to all selected sampling units from which households and/or individuals will be selected. Often these sampling units are villages, but could instead be city blocks, city districts, etc. depending on the sample design. Note: If household or individual selection is the <u>first or only</u> stage of sampling, it is not necessary to use Cluster IDs.	1-999
Household ID	All households to be visited should be assigned a unique ID. These numbers should be consecutive from 1 through the total number of households to be visited. This number can be assigned even before data collection begins because the total number of households to be visited should be known from the sample design. If no interview is conducted at a selected household, the Household ID assigned to it is simply not used.	1-99999
Participant ID	All participants should be assigned a unique ID. These need not be consecutive and a grouping by Cluster ID, where a sequence of participant IDs is associated with each Cluster ID, is strongly recommended (e.g. Participant IDs 101-120 are assigned to Cluster ID 1, Participant IDs 201-220 are assigned to Cluster ID 2, etc.).	1-999999999

Continued on next page

Preparing Data Collection Forms, Continued

Assigning Unique Identifiers (cont.)

The following three identifiers will also need to be assigned and made available to the data collection team as needed:

Variable	Description	Value Range
Data Collection Team ID	Each team of data collectors should be assigned a unique identifier and a record should be kept of which Interviewer IDs are associated with which data collection team.	A-Z
Technician ID	If Step 2 and/or Step 3 will be implemented by someone other than the interviewer (e.g. clinic staff), these individuals should be assigned a Technician ID.	1-999
Device ID	If Step 2 and/or Step 3 will be implemented, any equipment used for these Steps should be assigned a unique Device ID.	1-99

Note: The recommended value ranges indicate what is expected by the standard STEPS data entry templates. These templates will **not** allow for any mixing of alpha and numeric values (e.g. A21) in a single identifier.

Interview Tracking Form

All sites should use the Interview Tracking Form (see template in Part 6, Section 2) regardless of their sample design. This information is used for calculating the weights and response proportions for Step 1, Step 2, and Step 3 (if applicable).

Before data collection begins, Interview Tracking Forms should be completed for each Cluster and each interviewer who will conduct interviews in that Cluster.

Before data collection begins, the following should be completed on each Interview Tracking Form:

- Cluster ID
- Interviewer ID
- Household IDs
- Participant IDs.

Note: If household or individual selection is the first stage of sampling (i.e. Cluster IDs are not used), then prepare the Interview Tracking Forms for each interviewer without assigning Cluster IDs.

Continued on next page

Preparing Data Collection Forms, Continued

Kish Household Coversheet The Kish Household Coversheet (see a template in Part 6, Section 2) should be used when the data collection team needs to select participants randomly from each household. A Kish Household Coversheet must be prepared for every household to be visited.

Before data collection begins, the following should be completed on each Coversheet:

- Complete physical household address
- Household ID
- Participant ID.

Additionally, the column to be used in the Kish Selection Table can be circled as an added measure to help insure that the Kish Method is implemented correctly.

STEPS Instrument The following parts on the STEPS Instrument should be filled in prior to data collection:

- Cluster ID
 - Cluster name
 - Interviewer ID
 - Participant ID (on each page, first page twice).
-

Clinic Appointment Card The Clinic Appointment Card (see template in Part 6, Section 2) that serves for arranging appointments at the clinic for those selected for Step 3 should be partly filled in before the interviewers start data collection:

- Participant ID
- Centre name

should be filled in for those participants selected for Step 3.

Section 3: Preparing a STEPS Site





Overview

Introduction This section covers all the tasks that are needed to set up and prepare the STEPS site.

Intended audience This section is designed for use by those fulfilling the following roles:

- STEPS Site Coordinator
 - data collection team supervisors
 - data entry team supervisor.
-

Tasks and timeframes The chart below shows the main tasks and approximate timelines for setting up and preparing a STEPS site.

Task Name	Duration	Month 1	Month 2
Recruit staff	2 weeks		
Acquire equipment and supplies for data collection	2 weeks		
Acquire equipment for STEPS office	2 weeks		
Set up STEPS office	2 days		

In this section This section covers the following topics:

Topic	See Page
Recruiting Staff	2-3-2
Household Survey (Step 1 and 2)	2-3-4
Clinic Survey (Step 3 only)	2-3-7
Data Entry Office (Step 1, 2 and 3)	2-3-9

Recruiting Staff

Introduction The number and qualifications of staff will depend on the scope of the STEPS survey and the size of the sample as well as the type(s) of data to be collected, e.g. whether the site is implementing Step 1, 2 and 3, and if optional modules are added. See Part 1, Section 2 for further details on the roles and responsibilities described below.

Data collection team The core roles within the data collection team include some or all of the following:

- team supervisors
 - interviewers
 - survey clinic health professionals
 - laboratory technicians
 - administrative staff.
-

Data entry team The core roles within the data entry team include:

- team supervisors
 - data entry staff.
-

Data analyst The core tasks of the data analyst include:

- clean and prepare data for analysis
 - assist in creation of the Fact Sheet, Data Book, and STEPS Site Report.
-

Gender considerations For the data collection team, a mixture of staff of both sexes may be required for situations and communities where:

- there are strict rules about contact with members of the opposite sex
 - there is individual preference.
-

Language, ethnic and religious considerations For the data collection team, a mixture of staff who are fluent in several languages and represent varied cultural, ethnic and religious groups may be valuable.

Continued on next page

Recruiting Staff, Continued

Estimating numbers

The number and mix of staff requires careful calculation. For data collection, multiple teams should be recruited and trained to enable completion of interviews within the planned timeframe. All teams should have back-up staff available to cover for illness and other absences among members of the team.

See Part 2, Section 1 topic "Identifying Scope of STEPS Survey" for further details on estimating numbers of staff required.

Where to recruit people from

In many countries, recruitment is likely to be an informal process where data collection and data entry staff are 'seconded' from other duties within the Ministry of Health or other health authority responsible for undertaking the STEPS survey. In this situation, arrangements for their release and scheduled participation may need to be negotiated and explicitly agreed upon.

Where there is not sufficient available staff or specific skills are required (e.g. for data entry and data analysis) formal recruitment may be necessary.

Required timeframes

It is recommended that staff recruitment takes place over a period of 2-3 weeks if possible, so all may participate in initial training and build a good team structure quickly.

Household Survey (Step 1 and 2)

Introduction

Most sites will conduct Step 1 and 2 within households, although in rare cases, sites may choose to invite participants to attend a central location or a clinic.

General supplies for the household interviews

For the interviews you will need to prepare the following general supplies that the interviewer will have to take with him/her when collecting the data:

Materials	Quantity	Location of template
<ul style="list-style-type: none"> • district and area maps • household lists • name tags for interviewers • pens, pencils • clipboards 	1 of each	
• Interview Tracking Forms	2-3 per interviewer, depending on the number of interviews	Part 6, Section 2
• Notification Cards of WHO STEPS surveillance visit	1 per participant	Part 6, Section 2
• Kish Household Coversheets	1 per participant	Part 6, Section 2
• Participant Information Forms	1 per participant	Part 6, Section 2
• Consent Forms Step 2 and Step 3, if applicable	2 per participant, one stays with the participant	Part 6, Section 2
• STEPS Instruments	1 per participant	Part 5, Section 1
• Question-by-Question Guides	1 per interviewer	Part 5, Section 2
• show cards	1 per interviewer	Part 5, Section 3
• Participant Feedback Forms Step 2	1 per participant	Part 6, Section 2
• Body Mass Index Classification Charts	1 per interviewer	Part 6, Section 2
• if applicable: Clinic Appointment Cards	1 per participant	Part 6, Section 2
• if applicable: Fasting Instructions	1 per participant	Part 6, Section 2

Continued on next page

Household Survey (Step 1 and 2), Continued

General supplies for the household interviews (cont.)

The following documents should be prepared and partly filled in prior to the field work:

- Notification Cards of WHO STEPS surveillance visit (it is helpful for the interviewer if the contact details on the Notification Cards are already filled in prior to data collection);
- Kish Household Coversheets;
- Interview Tracking Forms;
- STEPS Instruments;
- if applicable: Clinic Appointment Cards.

Note: See Part 2, Section 2 for instructions on how to prepare the Kish Household Coversheets, Interview Tracking Forms, Clinic Appointment Cards and STEPS Instruments.

Equipment and supplies for Step 2 measurements

For Step 2, you will need the following specific equipment:

- adult, portable height-length measuring devices;
- weighing scales;
- stiff boards in case floor is uneven;
- constant tension tape measures (for example, Figure Finder);
- digital, automatic blood pressure monitors (OMRON device is recommended), complete with small, medium, large and extra large cuffs;
- spare batteries.

Note: Use of mercury sphygmomanometers is **not recommended** for general use but may be made available for use if the digital blood pressure monitor:

- is not functioning properly;
- needs calibration;
- if the largest cuff available on the digital device is too small for the participant.

Location for Step 2 measurements

Where STEPS is conducted entirely in a household setting, equipment and all supplies must be carried and set up as best as possible in each household. Each data collection team will carry the sets of equipment which are required.

If it is not possible to conduct the survey in each household, you may be able to identify a central location and schedule participants to visit at specified times.

Continued on next page

Household Survey (Step 1 and 2), Continued

Room setup for Step 2 measurements

Where a central location or public hall for taking Step 2 measurements is available, set up tables, chairs and equipment to optimize the flow of participants through the following steps:

Step	Action
1	Registration
2	Blood pressure measurement (and heart rate, if applicable)
3	Height measurement
4	Weight measurement
5	Waist circumference measurement
6	Hip circumference measurement, if applicable
7	Check out (to ensure all measures are complete and that participants are properly thanked for their participation before departure)

Note: Provide seating near where blood pressure will be measured to allow 15 minutes of relaxation before blood pressure measurement.

Other factors to consider

Some other factors to consider include:

Topic	Factors to consider
Equipment availability	Equipment necessary for collecting physical measurements should be readily available and in good condition to ensure results are as accurate as possible.
Lighting	Lighting needs to be adequate to read tape-measures, scales and blood pressure meters.
Weighing scales	Weighing scales need to be set up on a flat, hard surface. Some households may have uneven floors in which case an alternative location may need to be found or a rigid board should be placed under the scales.
Privacy	Areas used for taking measurements should be screened off or separated in some way to provide some privacy for participants.

Pre-survey site visits

It is advised that all proposed clusters/data collection sites are visited prior to conducting the survey.

This will allow a thorough understanding of operational issues that may impact the survey, and initiate the communication strategy with the communities and other local stakeholders.

Clinic Survey (Step 3 only)

Introduction

A clinic setting is necessary to take blood tests for biochemical measurements required in Step 3 of the STEPS survey.

Room and clinic location requirements

The following table lists the general requirements and set up considerations for the room and location chosen for taking biochemical measurements.

Item	Description
1	The room needs to be of adequate size to accommodate staff and the flow of the expected number of participants (and accompanying people).
2	Separate areas (if possible) for: <ul style="list-style-type: none">• registration• waiting• blood tests• checkout
3	Consider privacy requirements for taking blood tests
4	Provide hand washing and toilet facilities for participants and clinic staff
5	Clearly signpost the clinic
6	Ensure easy and adequate parking or transport provision for participants (if necessary)
7	Set up the room according to the sequence of tests

General equipment

General equipment required in the clinic is listed in the following table:

Material	Item
Stationery	<ul style="list-style-type: none">• pens• pencils• paper
Paperwork	<ul style="list-style-type: none">• Clinic Registration Form (partly filled in, see Part 2, Section 2) for each participant• STEPS Instrument (partly filled in, see Part 2, Section 2) for each participant• Participant Feedback Form Step 3 for each participant
Office equipment	<ul style="list-style-type: none">• filing systems• clipboard
Furniture	<ul style="list-style-type: none">• tables• chairs

Continued on next page

Clinic Survey (Step 3 only), Continued

Equipment and supplies

Different equipment is required depending on which type of chemistry has been selected for biochemical measurements. For further information about types of chemistry see Part 2, Section 1.

The table below provides a list of supplies required for the dry and wet chemistry methods.

Type	Supplies
Dry	<ul style="list-style-type: none">• batch of sufficient reagent test strips• lancet• lancet device cotton balls• alcohol swabs• gloves• disposable container
Wet	<ul style="list-style-type: none">• source of electric power• ice chests (and ice) for temporary storage• tourniquets• needles• syringes• primary and secondary specimen tubes• pipettes• gloves and possibly protective eyewear• centrifuge• facilities for safe disposal of used equipment particularly sharp needles and bloodied swabs etc.• transport of specimens

Data Entry Office (Step 1, 2 and 3)

Introduction The data entry office will need to accommodate the data entry team.

Room requirements The following table lists the general requirements and set up considerations for the space chosen for data entry work.

Item	Description
1	The room or rooms need to be of adequate size to accommodate all data entry staff and computers.
2	Set up tables to optimise physical work flows.
3	Create a pleasant environment for the team who will often sit for long periods.
4	Provide boxes or folders for instruments awaiting data entry.
5	Provide work-space for stacking papers at different stages of processing.
6	Provide temporary storage for individual instruments requiring problem resolution.
7	Set up a filing system for instruments once data entry and checking are completed.
8	Set up computers with good ergonomic positioning, to minimise reflections on screen, and to avoid build up of heat generated by machines.
9	Provide locked storage where instruments can be securely stored overnight and during weekends and holidays.

General equipment and supplies General office equipment and supplies required for the STEPS coordination and data entry office include:

- bench and table space
 - photocopier
 - shelving
 - filing cabinets or boxes
 - telephone
 - at least one computer with internet connection
 - office stationery supplies (paper, pens, envelopes, staplers etc).
-

Continued on next page

Data Entry Office (Step 1, 2 and 3), Continued

Computers

Where there is a choice of computer selection, refer to the list below for some general recommendations.

- Choose industry-standard computers and operating systems, i.e. IBM-compatible PCs running Microsoft Windows '98 or later.
 - Machines must have capability to transfer information (i.e. CD-writer, floppy disk, networked, or USB (flash disk) drive).
 - Purchase from reputable dealers.
 - If buying new, seek the highest specification machine(s) you can afford.
 - Speed of processing, memory capacity and hard disk space are important for data analysis but machine(s) with less memory may be adequate for data entry.
 - Desktop machines are usually cheaper and more easily maintained than portable machines.
 - Have at least two machines available to ensure backup in case of failure.
-

Printers

The quality of printer required is determined by the amount and type of printing the survey materials and data entry team needs. Use the following table to help determine what type of printer to use:

If the printer is used for...	Then choose a...
Producing lists, error checking and reporting progress.	Simple black-ink printer.
High-quality letters. Producing the main results, reports, tables and graphs.	Highly specified machine, possibly with colour capability.

Note: If purchasing a new printer, use reputable dealers and buy well-known, industry-standard machines and accessories.

Other equipment

Other equipment that may be needed depending on the location and facilities available include:

Purpose	Equipment options
Data backup	<ul style="list-style-type: none">• tape drivers, backup tapes• blank CDs• USB flash-stick• fireproof safe
Power supply	Uninterruptible power supply (UPS) machines

Continued on next page

Data Entry Office (Step 1, 2 and 3), Continued

Software

The following is a list of software that you will need to have setup on your office computers:

- Microsoft Office '98 or higher recommended for reports, correspondence and general word processing;
- virus scanning software (if connected to the internet and/or exchanging files outside the office);
- EpiData 3.1 (or later version, if available) for data entry.

For further information on data entry software, see Part 2, Section 4.

Section 4: Preparing the Data Entry Environment

Overview

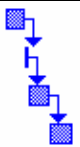
Introduction This section covers all the tasks that need to be conducted to setup, prepare and test the data files for the STEPS survey data entry.

Intended audience This section is designed for use by people who have been assigned the following roles:

- data entry supervisor
- data entry team
- STEPS Site Coordinator.

Note: These tasks may be commenced but not completed until the data entry team has been recruited.

Tasks and timeframes The chart below shows the main tasks and timelines covered in this section.

Task Name	Duration	Month 2
Map site-specific instrument	3 days	
Set up computer environment	1 day	
Modify data entry templates	3 day	
Test templates	4 days	

In this section This section covers the following topics:

Topic	See Page
Mapping the Site-Specific Instrument	2-4-3
Setting up the Computer Environment	2-4-4
EpiData	2-4-5
Data Entry Templates	2-4-8
Modifying the Templates	2-4-10
Additional Data Entry Files	2-4-14
Setting up the Data Entry Process	2-4-17
Testing	2-4-19
Documentation	2-4-22
File Security	2-4-23

Introduction

Overview of process

The table below lists each stage in the process of preparing the data entry environment.

Stage	Description
1	Mapping the site-specific instrument
2	Creating a Master computer
3	Accessing and installing EpiData
4	Installing and modifying the data entry templates
5	Creating STEPS survey data file folders
6	Installing the modified data entry templates on all other computers
7	Testing

Mapping the Site-Specific Instrument

Introduction The generic tools available for data entry and analysis require that the site-specific instrument be mapped to the generic instrument. In mapping the site instrument, one ensures that the code labeling each question in the site-specific instrument matches the code used for labeling that question in the generic instrument (e.g. C1, C2, T1, T2).

Purpose Mapping the site-specific instrument to the generic instrument will enable you to use:

- the data entry templates
- the data analysis programs
- the Fact Sheet Analysis Guide
- the Data Book.

Note: Once the site-specific instrument has been mapped, the data entry templates may need adjusting for site-specific response options and questions. Instructions for making these adjustments are covered later in this Section.

Available Materials There are two documents available for aiding you in mapping your site-specific instrument:

- Mapping and Transforming your Materials
- Mapped Instrument.

The first document, Mapping and Transforming your Materials, provides step-by-step instructions on how to map your site-specific instrument by using the second document, which provides a template with which to create your site-specific mapped instrument.

Both of these documents can be found on the STEPS website here:
<http://www.who.int/chp/steps/resources/mapped/en/index.html>

Site-specific Mapped Instrument Once the mapping is finished, you will have a complete site-specific mapped instrument that lists all questions in your site-specific instrument. For each question, the site-specific mapped instrument will indicate:

- the generic STEPS code that corresponds to the question (e.g. T1);
- the site-specific code used to label the question on the site-specific instrument;
- the possible response options for the question.

The site-specific mapped instrument will then serve as a useful aid in modifying the data entry templates. As it will also be helpful during data analysis, it should be made available to the data analyst.

Setting up the Computer Environment

Introduction

It is important to properly set up your computer environment prior to working with any data files.

Create Master computer and label others

Designate and label one of the computers in the STEPS Office as the Master computer. This computer will be used to install, modify and test the data entry templates prior to installing them on the other computers.

Label all the other computers (e.g. A, B, C, D, etc.).

Setting up the data entry computers

Follow the steps below to create appropriate folders on the data entry computers for all EpiData surveillance files:

Step	Action	Recommended Folder Name
1	Create a primary folder (directory) for all your STEPS files, including: <ul style="list-style-type: none">• data• code• documents• other files.	C:\STEPS
2	Record the address of the folder so it can be entered during the set-up process when prompted.	
3	Create a backup folder in a different location than the primary folder.	<ul style="list-style-type: none">• D:\STEPS (or similar, if you have multiple drives or your disk is partitioned)• C:\BackupSTEPS (if you only have access to one drive)
4	Create a sub-folder under the STEPS primary folder to contain your data files.	C:\STEPS\data
5	Create a sub-folder under STEPS\data to contain data entry reports.	C:\STEPS\data\reports
6	Create a sub-folder under STEPS\data to contain office tracking information.	C:\STEPS\data\office

EpiData

Introduction

To enter the STEPS survey data, the STEPS team recommends and supports using EpiData 3.1. EpiData 3.1 is a purpose-built, free, public-domain software package that allows users to:

- capture the survey data
 - verify data entry accuracy (93).
-

Rationale

The decision for choosing EpiData was made in light of its advantages, some of which are listed below.

- Windows-based and compatible with other software;
 - widely used;
 - makes compact and easily modifiable data files;
 - checks for valid ranges during data entry, but permits values beyond ordinary ranges;
 - allows double data entry and data correction;
 - files exportable to 6 different file types.
-

Accessing EpiData

The current release of EpiData is available on the STEPS CD as well as on the STEPS website here:

<http://www.who.int/chp/steps/resources/EpiData/en/index.html> .

Additionally, EpiData can be downloaded and installed directly from the EpiData website: www.epidata.dk .

The table below provides instructions on how to get the EpiData installation file onto your computer.

Source	Instructions	
STEPS website	Step	Action
	1	Connect to the internet and go to: http://www.who.int/chp/steps/resources/EpiData .
	2	Click on the link labeled "Download EpiData Software".
	3	Save the installation file, "setup_epidata.exe", to your desktop.

Continued on next page

EpiData, Continued

Accessing EpiData (cont.)

Source	Instructions	
STEPS CD	Step	Action
	1	Insert the CD into the CD-ROM drive and wait for the CD to launch in your internet browser. If the CD does not launch automatically, go to the list of all files on the CD and open the file "start.html".
	2	Click on the link labeled "Data Entry Tools and Software" in the left-hand column of the screen.
	3	Click on the link labeled "EpiData Software".
	4	Click on the link labeled "Download EpiData Software".
	5	Save the installation file, "setup_epidata.exe", to your desktop.

Installation

Once the EpiData installation file has been downloaded to your computer, follow the steps below to install EpiData:

Step	Action
1	Go to your desktop and click on the file "setup_epidata.exe".
2	Click "Yes" on the dialog box that says you will install the program. Click "Next" on the welcome screen to continue installation.
3	Read the licensing screen and click "I accept the agreement" and click "Next".
4	An installation program will start, when prompted to select a destination directory make sure the location is "C:\Program Files\EpiData". Click "Next".
5	Click "Don't create a start menu folder". Click "Next".
6	Select "Create a Desktop icon" and "Automatic field naming" from the Select Additional Tasks page. Click "Next".
7	Review the information on the Ready to install screen. If the information is correct click "Install" if it is incorrect use the "Back" button to correct the information.
8	Once you have confirmed that EpiData is properly working on your machine, you can delete the file "setup_epidata.exe" from your desktop.

Note: EpiData will need to be installed on all computers that will be used for data entry.

Continued on next page

EpiData, Continued

Training

EpiData training materials are available in Part 3, Section 5 of this manual.

Additionally, further reading about data entry in EpiData may be found at the EpiData website: www.epidata.dk.

Of particular interest is the EpiData extended help file, which is available on the EpiData website here: <http://www.epidata.dk/documentation.php>.

Software support

WHO provides some support for EpiData and can assist in the modification of the data entry templates. If you use software other than EpiData, you are responsible for creating your own data entry files and obtaining suitable support.

Data Entry Templates

Introduction

Standard STEPS templates have been developed to enter survey data from completed instruments. These must be reviewed and updated as needed to make sure they match your site-specific instrument.

Installing EpiData templates from the web

Follow the steps below to install the EpiData templates from the STEPS website.

Step	Action
1	Connect to the internet and go to: http://www.who.int/chp/steps/resources/EpiData .
2	Under the section header "Data Entry Templates", select and download the template zip file that matches your instrument.
3	Save the file in C:\STEPS\data.
4	Open the C:\STEPS\data folder and double click the zip file you have downloaded.
5	The zip file will open up and display several folders. Highlight these folders and copy them. Close the zip folder and paste the folders directly into C:\STEPS\data.

Note: It is recommended that you only install the templates to the Master computer initially and then copy them from the Master computer to the other computers only once any needed modifications have been done.

Templates

There are 3 generic templates that have been developed for EpiData. Each is located in a separate folder of the same name. The table below lists and describes the purpose of each of these templates.

Template	To contain
Consent	Personal information from the instrument (if to be saved electronically).
Survey	Location and date of interview and main instrument data.
Biochemical	Step 3 results, if these are recorded on a form separate from the instrument.

File types

For each template, there are several EpiData files that combine together to make a functioning data entry template. The table below describes the 3 key files for each template:

Extension	Purpose
.rec	Used to enter data. Entered records are saved in this file.
.qes	Used to create the data entry interface. The .rec file is generated from this file.
.chk	Contains tests for out of range values and instructions for skipping questions that are not applicable.

Continued on next page

Data Entry Templates, Continued

Consent template (optional)

The optional consent template collects the confidential data from the lower half of the Survey Information page of the instrument. These data should not be stored with the information entered in any other template and should not be used during data analysis. It may be useful to store this information electronically if:

- participants need to be contacted after the interview;
 - quality control procedures require follow-up contacts;
 - participants are advised to visit their clinic or physician where biochemical results indicate medical attention (if appropriate).
-

Survey template

This template is for recording all information pertaining to the location and date of the interview as well as the main STEPS core, expanded and optional data. Each record in this database is uniquely identified by the Participant ID. The final structure of the template should match exactly the site-specific instrument. Data collected includes:

- Cluster Name
 - Cluster ID
 - Interviewer ID
 - date of completion of the instrument
 - core questions and measures for Step 1 and 2 (and may include Step 3)
 - expanded and/or optional questions.
-

Biochemical template

This template applies only to sites conducting Step 3 and who have recorded Step 3 data on a separate form.

Modifying the Templates

Introduction

If you have made any modifications to the generic STEPS Instrument in creating your site-specific instrument, you will probably need to modify the data entry template to reflect these changes. Possible changes to the generic STEPS Instrument that require modification of the data entry templates include:

- changing the response options for a question
 - changing the wording of a question
 - adding new questions
 - excluding any core or expanded questions.
-

Role and responsibility

The data entry supervisor should be responsible for modifying the templates. The WHO Geneva STEPS team will also help in template modification upon request.

Preparing for template modification

Before making any modifications to the questionnaire, you must first do a thorough assessment of the differences between your site-specific instrument and the generic STEPS Instrument. If any differences are not accounted for, this may cause serious problems during data entry. The table below provides a series of steps to follow to ensure a thorough preparation for template modification.

Step	Action
1	Ensure that you have a complete and correct copy of the site-specific mapped instrument as well as a copy of the site-specific instrument.
2	Read carefully through the site-specific mapped instrument, highlighting all differences between the site-specific instrument and the generic STEPS Instrument.
3	Using the site-specific mapped instrument as a guide, read carefully through the site-specific instrument marking all questions that have been highlighted in the site-specific mapped instrument.

Modifying the templates

To make the actual changes to the templates, you will need to modify and update the appropriate EpiData files. The table below shows the type of modification and what corresponding data files need to be updated.

To	Update the following files
Alter response options for a question	.chk
Alter the wording of a standard question	.ges .rec
Add an optional question	.ges .rec .chk
Hide an unused question	.chk

Continued on next page

Modifying the Templates, Continued

Altering response options

Follow the steps below to change the wording of a response option or to add additional response options:

Step	Action
1	Open EpiData.
2	Click on "3.Checks" at the top of the screen.
3	Select the .rec file you wish to modify and click "Open".
4	Click on the yellow box to the right of the question you wish to edit.
5	When the box turns blue, click on "Edit" in the small dialogue window.
6	Locate the text that needs to be changed (e.g. "locally defined 1") and replace it with the new text, making sure to surround the new text with quotation marks. If additional response options are needed, add them below the list of existing options making sure to follow the same format.
7	When finished making changes, click on "Accept and Close" at the top of the window. The .chk file will be updated to reflect the changes made.

Alter question wording

Follow the steps below to change the wording of a question:

Step	Action
1	Open EpiData.
2	Click on "1.Define Data" at the top of the screen.
3	Select the .ges file you wish to modify and click "Open".
4	Alter the wording of the question as needed (it is just like modifying a text document).
5	When finished, click on "2. Make Data File" and follow the prompts to recreate the .rec file.

Continued on next page

Modifying the Templates, Continued

Adding a new question

Follow the steps below to add a new question to the data entry template:

Step	Action
1	Open EpiData.
2	Click on "1.Define Data" at the top of the screen.
3	<p>Find the location in the file where the new question should be entered. Create a new line in which to insert the question and type:</p> <ul style="list-style-type: none"> • the code for the question (e.g. X1); • the text for the question; • the response type for the question (e.g. ## for numeric, ___ for text). <p>These three items should be delimited by at least one space. For help with the response type, the Field Pick List dialogue can be opened by clicking on the 2nd icon from the right at the top of the screen. Note that the number of characters entered for the response type reflects exactly the number of characters that may be entered in that field during data entry.</p>
4	Click on "2. Make Data File" at the top of the screen.
5	Add checks for the new question by clicking on "3. Checks" at the top of the screen and selecting the .rec file that was just modified. Click on the response field for the new question to add/edit checks. The following page contains some sample check code.

Hiding an unused question

Unused questions should **not** be deleted from the template. Instead, they should be hidden. Follow the steps below to hide any unused questions:

Step	Action
1	Open EpiData.
2	Select "Open" from the "File" menu at the top of the screen and select the .chk file you wish to modify.
3	Find the text "BEFORE FILE" in the .chk file.
4	<p>In this section of the file, list all the questions you wish to hide on a separate line, each preceded by the word "HIDE".</p> <p>Example (3 questions hidden):</p> <pre>BEFORE FILE DEFINE varEntryBegun #####.##### HIDE C5 HIDE C6 HIDE C10 END</pre>
5	Save the .chk file by going to File/Save in the menu at the top of the screen.

Continued on next page

Modifying the Templates, Continued

What not to modify

You must never:

- change field names (e.g. C1, C2) for existing questions
- delete questions from the data entry screen (.qes file).

If these items are modified, the template will not work.

Sample check code

Samples of the three types of check code are provided in the table below.

Type of check	Sample code	Function
Create value labels.	<pre>C1 COMMENT LEGAL 1 Male 2 Female END TYPE comment END</pre>	Creates a list of possible responses with their respective labels (e.g. yes/no).
Provide range checking of values and avoid missing data.	<pre>c3 AFTER ENTRY IF (c3=.) THEN HELP "An age must be entered." GOTO c3 EXIT ENDIF IF ((c3<15) OR (c3>74)) AND (c3<>77) THEN GOTO c3 EXIT ENDIF END END</pre>	Ensures a value is entered within a range of values that are acceptable for data entry. This helps ensure more accurate data entry (e.g. this example allows values from 15 to 74, or 77 for don't know).
Provide skipping of questions.	<pre>T1 AFTER ENTRY IF T1=2 THEN GOTO T6 ENDIF END END</pre>	Mimics the skip pattern on the instrument. It will take the data entry person directly to the next applicable question.

Check your work

After each template is modified, it is important to make sure that the template is still in working order and ready for testing. To perform a quick check of your work, open the modified template by double-clicking on the .rec file. Scan through the data entry screen to ensure it appears as desired and ensure that appropriate data can be entered into the modified/added fields.

Additional Data Entry Files

Introduction

In addition to the data entry templates, there are 3 additional files that should be obtained from the STEPS website or CD, these are:

- interview_tracking_form.xls
 - data_entry_log.xls
 - data_entry_tracking_from.doc.
-

Interview Tracking Form

The Interview Tracking Forms should be completed by the interviewers during data collection. The data from these forms should be entered using the Excel file interview_tracking_form.xls. It is recommended that all Interview Tracking Forms be entered on the Master computer.

The interview_tracking_form.xls file is available from STEPS CD and the STEPS website. It will only need to be downloaded to the Master computer.

Download from ...	Instructions	
CD	Step	Action
	1	Click on "Data Entry Tools and Software" in the left-hand column of the home screen.
	2	Click "Interview Tracking Spreadsheet".
	3	Save the file in C:\STEPS\data.
STEPS website	Step	Action
	1	Go to the resources section of the STEPS website: www.who.int/chp/steps/resources .
	2	Click "Interview Tracking Spreadsheet".
	3	Save the file in C:\STEPS\data.

Data Entry Log Excel file

The data_entry_log.xls file should be completed by the data entry supervisor. The file enables the supervisor to record for each completed instrument received:

- Participant ID
 - date received
 - to whom the 1st keying was assigned
 - to whom the 2nd keying was assigned
 - problems, solutions, and additional notes
-

Continued on next page

Additional Data Entry Files, Continued

Data Entry Log Excel file (cont.) The data_entry_log.xls file is available on the STEPS CD and the STEPS website. It will only need to be downloaded to the Master computer.

Download from ...	Instructions	
CD	Step	Action
	1	Click on "Data Entry Tools and Software" in the left-hand column of the home screen.
	2	Click "Data Entry Log".
	3	Save the file in C:\STEPS\office.
STEPS website	Step	Action
	1	Go to the resources section of the STEPS website: www.who.int/chp/steps/resources .
	2	Click "Data Entry Log".
	3	Save the file in C:\STEPS\office.

Data Entry Tracking Form The Data Entry Tracking Form is available in both Word and Excel. Each computer should have one Data Entry Tracking Form which will be used by data entry staff to record for each instrument entered on that computer:

- Participant ID
- date of 1st keying
- date of 2nd keying
- errors found on instrument
- supervisor's decision to handle error (where applicable).

Continued on next page

Additional Data Entry Files, Continued

Data Entry Tracking Form (cont.)

The Data Entry Tracking Form is available on the STEPS CD and the STEPS website. One copy should be downloaded onto each data entry machine.

Download from ...	Action	
CD	Step	Action
	1	Click on "Data Entry Tools and Software" in the left-hand column of the home screen.
	2	Click "Data Entry Tracking Form (Excel version)" or "Data Entry Tracking Form (Word version)".
	3	Save the file in C:\STEPS\office.
STEPS website	Step	Action
	1	Go to the resources section of the STEPS website: www.who.int/chp/steps/resources .
	2	Click "Data Entry Tracking Form (Excel version)" or "Data Entry Tracking Form (Word version)".
	3	Save the file in C:\STEPS\office.

Setting up the Data Entry Process

Introduction

Prior to receiving completed instruments in the STEPS office for data entry, you will need to set up a standard working method to ensure accurate and efficient handling of survey material and data entry.

Working method

Create a standard working method that includes the following elements:

- labels for computers being used for data entry
 - boxes or folders for each computer to store instruments and tracking forms
 - coversheets for computer-specific folders/boxes
 - data entry guidelines and rules (protocols)
 - a Data Entry Tracking Form specific to each data entry computer
 - data entry staff assigned to specific data entry computers.
-

Labeling computers

Where there is more than one computer being used for data entry, you will need to label each machine so you can enter and track specific information as shown in the table below.

Computer	To enter
Master	<ul style="list-style-type: none">• Instrument responses to Step 1, Step 2 and Step 3 (where appropriate)• Biochemical (if Step 3 not recorded on instrument)• Tracking information (Interview Tracking Forms)
A, B, C etc.	<ul style="list-style-type: none">• Instrument responses to Step 1, Step 2 and Step 3 (where appropriate)• Biochemical (if Step 3 not recorded on instrument)

Continued on next page

Setting up the Data Entry Process, Continued

Filing Process Establish a system of boxes or folders to store the hard copies of the instruments that have been or will be entered on each computer. Label these with the coversheet (the coversheet template is provided in Part 6, Section 2). The table below describes the 3 folders that should be made for each data entry computer.

Folder	For instrument data that...	Folder name
1	Is not yet entered	1 st key
2	Has first keying complete	2 nd key
3	Has second keying complete	Completed

Note: If the Consent or Biochemical data entry templates will be used, a separate set of folders will be needed for each of these templates. If a machine is being used to enter data into more than one template, a set of folders should be created for each template for that machine.

Protocols Create data entry protocols to cover each of the key stages in the data entry process, including:

Process	Guidelines or rules required to
Handling incoming instruments	Specify how to sort, label and handle the completed incoming instruments from the data collection team.
Data entry	Specify how data entry staff will perform the data entry process and what they should do when they find unexpected or ambiguous data.
Marking and filing	Ensure any paper can be easily located at any time, and all instruments and forms show on them their stage of processing.
Handling uncertain data	Obtain a supervisor's ruling on uncertain data and a method for documenting what decisions are made.
Documentation	Ensure an audit trail of all completed and altered records.

Data entry staff You should permanently assign data entry staff to work at two specific computers for the entire data entry process. Each staff member will be responsible for first keying the instruments on one computer and for second keying the instruments on the other computer.

Testing

Introduction

Once the templates have been modified, the data entry screen and all data entry systems and processes must be thoroughly tested to identify and correct any problems prior to data entry.

The two test phases are:

- primary testing
 - pilot testing of all data entry processes.
-

Overview

The table below gives an overview of the testing process.

Type of test	Who should be involved	Preparation	Time Frame
Primary test	Data entry supervisor or person responsible for modifying the templates	<ul style="list-style-type: none">• Complete all planned modifications to data entry templates on the Master computer.• Have a copy of the site-specific instrument and site-specific mapped instrument ready for use during testing.	Half a day
Pilot test data entry processes	Data entry staff (and/or members of the data collection team if necessary) and the data entry supervisor	<ul style="list-style-type: none">• Complete primary test of data entry templates.• Copy the modified data entry templates from the Master computer to the C:\STEPS\data folder on all other data entry computers.	1-2 days

Continued on next page

Testing, Continued

Primary test

Follow the steps below to run the primary test. This should be done on the Master computer and by the same person who modified the templates.

Step	Action										
1	Using the finalized site-specific instrument, create 8-12 completed "interviews". <ul style="list-style-type: none"> • Use different coloured paper or otherwise distinguish between these test forms and real ones by labeling them as test. • Make them straightforward, correct and clear, but with a variety of "participants" (e.g. smokers & non-smokers, active & sedentary). 										
2	Create a new folder titled "C:\TestSTEPS" on the Master computer.										
3	Copy the entire STEPS folder and paste it into the new test folder.										
4	Use the "C:\TestSTEPS" for the testing phase.										
5	Run an initial test to check the templates. <table border="1" data-bbox="560 943 1428 1137"> <thead> <tr> <th>Step</th> <th>Action</th> </tr> </thead> <tbody> <tr> <td>5.1</td> <td>Open EpiData.</td> </tr> <tr> <td>5.2</td> <td>Click "4. Enter Data".</td> </tr> <tr> <td>5.3</td> <td>Select the template to test.</td> </tr> <tr> <td>5.4</td> <td>Enter the 8-12 selected "interviews".</td> </tr> </tbody> </table>	Step	Action	5.1	Open EpiData.	5.2	Click "4. Enter Data".	5.3	Select the template to test.	5.4	Enter the 8-12 selected "interviews".
Step	Action										
5.1	Open EpiData.										
5.2	Click "4. Enter Data".										
5.3	Select the template to test.										
5.4	Enter the 8-12 selected "interviews".										
6	Update the templates in "C:\STEPS" with corrections as needed and repeat steps 3-5 above.										
7	Once all templates have been tested and no further corrections are needed, proceed with preparations for the pilot test.										

Continued on next page

Testing, Continued

Pilot test

After providing basic data entry training to your data entry team, a pilot test should be done to thoroughly test the modified templates and the entire data entry process. The data entry training and pilot test are covered in more detail in Part 3, Section 5. The pilot test entails:

- entering test data into the interview_tracking_form.xls file
 - entering test data into the EpiData data entry templates
 - testing all logging and sorting processes
 - testing all error correction systems.
-

Finalizing the data entry process

Once the pilot test has been completed and any problems discovered have been fixed, copy all finalized templates to all computers.

Step	Action
1	On the Master computer copy the folder "C:\STEPS" and all its contents onto a CD or USB stick (flash disk).
2	Go to Machine A.
3	Open the C drive.
4	Copy folder from the CD or USB stick onto the C drive, replacing the C:\STEPS folder that exists there already.
5	Create a backup folder in a different location than the primary folder (we recommend D:\STEPS, if possible).
6	Repeat steps 2-5 until all data entry computer installations are complete.

Documentation

Introduction Documentation is essential for an efficient and effective STEPS survey.

Documenting data entry The data entry process must be documented to ensure:

- standardization of processes and procedures among all data entry team members;
- non-reliance on certain individuals to provide key information;
- easy access to essential information, regardless of absence;
- data entry and data analysis can be done when the person who created the database is not available;
- survey data comparisons are possible in the future.

The Data Entry Log and Data Entry Tracking Form (see page 2-4-14 and Part 3, Section 5) should be used to assist in documenting the data entry process. They are available in Excel format but should be printed regularly for backup purposes.

Other documentation requirements All survey files and resources must be:

- stored systematically (both paper and electronic)
- fully documented continuously.

Don't plan to come back later to annotate: make it a habit to place comments on your files as you work.

File Security

Introduction The information collected by STEPS needs to be kept in a secure location. This applies both to the paper copies and electronic information.

Paper copies Paper copies should be locked up every night in a secure location.

Electronic information The computers that are used for data entry and analysis need to be located in a secure location. If the computers are in a locked location it is not necessary to place a password on the machines.

If computers are in a shared space and cannot be locked up at night, it is best to place a logon password on each machine.

Note: If you decide to place passwords on the machines please make sure the data entry supervisor has a complete list of passwords for each machine.

Backup At the end of **each day** of data entry you must backup all your data files. This is to avoid data loss.

Further details on backing up the data are provided in Part 3, Section 5.

Section 5: Preparing the Data Analysis Environment

Overview

Introduction This section covers all the tasks that need to be conducted to setup and prepare for the analysis of the STEPS survey data.

Intended audience This section is designed for use by people who have been assigned the following roles:

- data analyst
 - statistical adviser
 - STEPS Site Coordinator.
-

Timeframe The set up of the data analysis environment can be done within one day. However, analysis of the survey data cannot proceed until data entry has been completed and the data entry supervisor has provided a finalized data set to the data analyst.

In this section This section covers the following topics:

Topic	See Page
Epi Info	2-5-2
Setting Up the Computer Environment	2-5-4
Preparing the STEPS Data for Analysis	2-5-5
Data Analysis Programs	2-5-7

Epi Info

Introduction

To check and analyse the STEPS surveillance data, the STEPS team recommends and supports using Epi Info. Epi Info is a purpose-built, free, public-domain software package that allows users to:

- check the survey data for outliers and inconsistent data
 - conduct a descriptive analysis of survey data
 - easily generate output files from the analysis (94).
-

Rationale

The decision for choosing Epi Info was made in light of its advantages, including:

- Windows-based
 - supported by developers
 - has data analysis capability in line with STEPS requirements
 - can appropriately adjust for complex sample designs.
-

Accessing Epi Info

The current release of Epi Info is available on the STEPS CD as well as on the STEPS website. Additionally, Epi Info can be downloaded and installed directly from the Epi Info website (<http://cdc.gov/epiinfo/>).

The table below provides instructions on how to get the Epi Info installation file onto your computer.

Source	Instructions	
STEPS website	Step	Action
	1	Connect to the internet and go to: http://www.who.int/chp/steps/resources/EpiInfo .
	2	Click on the link labeled "Epi Info 3.4.3".
	3	Save the installation file, "setup_epiinfo.exe", to your desktop.
STEPS CD	Step	Action
	1	Insert the CD into the CD-ROM drive and wait for the CD to launch in your internet browser. If the CD does not launch automatically, go to the list of all files on the CD and open the file "start.html".
	2	Click on the link labeled "Data Analysis Tools and Software" in the left-hand column of the screen.
	3	Click on the link labeled "Epi Info Software".
	4	Click on the link labeled "Download Epi Info 3.4.3".
5	Save the installation file, "setup_epiinfo.exe", to your desktop.	

Continued on next page

Epi Info, Continued

Installing Epi Info

Once the Epi Info installation file (setup_epiinfo.exe) is on your desktop, follow the instructions below to install Epi Info on your machine.

Step	Action
1	Double click on the Epi Info installation file on your desktop.
2	Click "Next" on the Welcome to Epi Info 3.4.3 screen.
3	Click "Next" on the Destination folder screen, you should use the default C:\Epi_Info.
4	Click "Next" on the selected features screen.
5	Click "Next" on the ready to install screen.
6	After ensuring the installation is successful, it is safe to delete the Epi Info installation file from your desktop.

Learning Epi Info

The table below lists several resources available to help you learn to use Epi Info.

Resource	Content	Location
Data Analysis Guide	A brief summary of useful commands in the Analysis module of Epi Info.	STEPS Manual, Part 3, Section 6
Epi Info Training Guide and supporting materials	An in-depth Epi Info training manual with numerous examples and related training materials.	STEPS CD and STEPS website: http://www.who.int/chp/steps/resources/EpiInfoTraining/en/index.html
Epi Info Tutorials	Tutorials provided by CDC and links to other learning resources.	CDC website: http://cdc.gov/epiinfo/tutorials.htm

Software support

WHO provides some support for Epi Info and can provide assistance and training as needed. If you use software other than Epi Info, WHO will only be able to provide little, if any, support.

Setting Up the Computer Environment

Introduction

It is critical to properly set up your computer environment prior to attempting any analysis of your STEPS data in Epi Info.

Preparing Folders

The following folders must be created on the computer used for data analysis. The names listed for these folders in the table below are not optional. The folders **must** be named using the following naming conventions in order to use the analysis code provided by the STEPS team.

Step	Action	Required Folder Name
1	Create a primary folder (directory) for all your STEPS files, including: <ul style="list-style-type: none">• data• analysis output.	C:\STEPS
2	Create a backup folder in a different location than the primary folder to store a backup copy of your STEPS dataset.	<ul style="list-style-type: none">• D:\STEPS (or similar, if you have multiple drives or your disk is partitioned)• C:\BackupSTEPS (if you only have access to one drive)
3	Create a sub-folder under the STEPS primary folder to contain your STEPS dataset ready for analysis in Epi Info.	C:\STEPS\EpiInfo
4	Create a sub-folder under STEPS\EpiInfo to contain analysis output.	C:\STEPS\EpiInfo\ Output Tables
5	Create a folder under the STEPS primary folder for any additional data files (e.g. interview_tracking_form.xls).	C:\STEPS\data

Preparing the STEPS Data for Analysis

Introduction

Upon completion of the data entry process, the data entry supervisor should have a single data file containing all STEPS survey data as well as a completed interview_tracking_form.xls file containing all interview tracking information. Both of these files must be prepared for analysis by the data analyst.

Note: The instructions that follow assume that EpiData was used for data entry and that the final version of the complete STEPS survey data was exported from EpiData into a .dbf file. If this was not the case for your data, please contact the STEPS team for assistance.

Overview

In order to analyse your STEPS data in Epi Info using the generic programs available from the STEPS team, the data must be transported to a Microsoft Access database containing your survey data and the generic programs, each in a separate table. The instructions below describe one means to move your data into the Access database.

Preparing the STEPS survey data

Follow the steps below to prepare your STEPS data for analysis in Epi Info.

Step	Action
1	Rename the .dbf file containing your survey data to "STEPS.dbf" (right-click on the file and select "Rename" to edit the file name).
2	Open the .dbf file in Access. Note that your dataset is listed in the Database window as a linked table (indicated by a blue arrow followed by the letters dB).
3	Copy the dataset to a local table by right-clicking on the dataset and selecting "Copy". Then right-click on any white space in the Database window and select "Paste".
4	In the Paste Table As dialog window, type "MasterDataSet" in the space for Table Name and select "Structure and Data (Local Table)" from the Paste Options.
5	After ensuring the name is spelt correctly and the correct option is selected, click "OK".
6	You will now see the local data table, MasterDataSet, listed in your Database window. You may open it to have a look at your data.
7	Exit from Access. You will see that you now have a STEPS.mdb file listed along with your STEPS.dbf file.
8	Move the STEPS.mdb file to the C:\STEPS\EpiInfo folder.

Continued on next page

Preparing the STEPS Data for Analysis, Continued

Preparing the interview tracking form Excel file

Follow the steps below to prepare the completed interview_tracking_form.xls for analysis in Epi Info.

Step	Action						
1	Save the interview_tracking_form.xls file to the C:\STEPS\EpiInfo\data folder.						
2	Open the interview_tracking_form.xls file and click on the worksheet entitled "enter information". Check to see that this table has been filled in. If not, stop and contact the data entry supervisor.						
3	After ensuring that the "enter information" worksheet has been completed, check to see if the worksheet "EpiInfo" exists. Refer to the table below to decide how to proceed: <table border="1" data-bbox="555 846 1396 1182"> <thead> <tr> <th>If ...</th> <th>Then ...</th> </tr> </thead> <tbody> <tr> <td>"EpiInfo" worksheet exists</td> <td>Exit Excel.</td> </tr> <tr> <td>"EpiInfo" worksheet does not exist</td> <td>Go to the "Instructions" worksheet in the file and click on the button labeled "Format for Epi Info". After doing so, an "EpiInfo" worksheet should appear. You can now exit Excel.</td> </tr> </tbody> </table>	If ...	Then ...	"EpiInfo" worksheet exists	Exit Excel.	"EpiInfo" worksheet does not exist	Go to the "Instructions" worksheet in the file and click on the button labeled "Format for Epi Info". After doing so, an "EpiInfo" worksheet should appear. You can now exit Excel.
If ...	Then ...						
"EpiInfo" worksheet exists	Exit Excel.						
"EpiInfo" worksheet does not exist	Go to the "Instructions" worksheet in the file and click on the button labeled "Format for Epi Info". After doing so, an "EpiInfo" worksheet should appear. You can now exit Excel.						

Data Analysis Programs

Introduction

Generic data analysis programs have been prepared by the STEPS team to complete basic descriptive analyses of STEPS survey data in Epi Info. These programs can be used to create all the output needed to complete the Fact Sheet and Data Book.

Accessing the Data Analysis Programs

The data analysis programs are available on the STEPS CD as well as on the STEPS website. As these are occasionally updated to meet country needs, it is best to check the updates page on the STEPS website (<http://www.who.int/chp/steps/resources/updates/en/index.html>) to see if a more recent version is available.

The table below provides instructions on how to get the analysis programs onto your computer.

Source	Instructions	
STEPS website	Step	Action
	1	Connect to the internet and go to: http://www.who.int/chp/steps/resources/database/en/index.html .
	2	Click on the link labeled "Epi Info Analysis Programs".
	3	Save the zip file, "Epi_Info_Analysis_Programs.zip", to your desktop.
	4	Open the zip file by double-clicking on it. Copy the Access file, Epi_Info_Analysis_Programs.mdb, to your desktop.
STEPS CD	Step	Action
	1	Insert the CD into the CD-ROM drive and wait for the CD to launch in your internet browser. If the CD does not launch automatically, go to the list of all files on the CD and open the file "start.html".
	2	Click on the link labeled "Data Analysis Tools and Software" in the left-hand column of the screen.
	3	Click on the link labeled "Epi Info Analysis Programs".
	4	Save the Access file, Epi_Info_Analysis_Programs.mdb to your desktop.

Continued on next page

Data Analysis Programs, Continued

Attaching the Data Analysis Programs to your Data

Once the analysis programs have been downloaded to your desktop, you will need to attach the programs to your STEPS.mdb file containing your STEPS data.

Step	Action
1	Open the Epi_Info_Analysis_Programs.mdb file and right click on the table "Programs".
2	Select "Export" and in the "Export Table 'Programs' To" window type "C:\STEPS\EpiInfo\STEPS.mdb" in the "File Name" field and click "Export".
3	Select the "Definition and Data" option in the dialogue window and click "OK".
4	If you are updating an older version of the programs, a dialogue window will pop up that asks if you want to replace the "Programs" table that already exists. Click "Yes".
