

# Part 4: Conducting the Survey, Data Entry, Data Analysis and Reporting and Disseminating Results

## Overview

---

**In this Part**

This Part covers the following topics

<b>Topic</b>	<b>See Page</b>
Section 1: Data Collection	4-1-1
Section 2: Data Entry	4-2-1
Section 3: Data Analysis	4-3-1
Section 4: Reporting and Disseminating Results	4-4-1
Section 5: Archiving	4-5-1

---



# Section 1: Data Collection

## Overview

---

**Introduction** This section covers all the tasks that need to be undertaken to:

- supervise data collection
  - approach the selected households
  - inform participants and obtain consent
  - track participation
  - conduct the interviews
  - take the measurements
  - record the data collected
  - fill in the Participant Feedback Form
  - schedule clinic visits for Step 3 measurements.
- 

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

- data collection team supervisor
  - data collection staff
  - STEPS Site Coordinator.
- 

**Timeframes** Data collection takes approximately 8-12 weeks. This depends, however, on the number of staff available as well as on the logistics in a country / site.

---

**In this section** This section covers the following topics:

<b>Topic</b>	<b>See Page</b>
Supervising Data Collection	4-1-2
Data Collection Process and Interviewer Tasks	4-1-5
Approaching Selected Households and Participants	4-1-7
Informing Participants	4-1-9
Obtaining Consent	4-1-11
Completing the Interview Tracking Form	4-1-12
Recording Information	4-1-14
Completing the Participant Feedback Form	4-1-16
Scheduling Clinic Visits for Step 3 Measurements	4-1-17

---

## Supervising Data Collection

---

**Introduction** Members of the data collection team may have different levels of skills, experience and varying strengths and abilities. To ensure high standards of data collection, appointing one or more persons to lead and supervise the data collection team(s) is necessary.

---

**Core tasks** The core tasks of a data collection team supervisor are provided in the checklist below. General roles are identified in Part 1, Section 2.

Tasks	Description	✓
1	Train data collection team staff.	
2	Obtain lists of the selected sample and maps for each area and assign interview teams to locations.	
3	Contact local authorities.	
4	Make travel arrangements for data collection teams.	
5	Obtain and prepare data collection forms and copies of the questionnaire.	
6	Distribute forms and supplies to interviewers.	
7	Track interviews.	
8	Supervise data collection.	
9	Supervise human resources.	
10	Provide progress updates to the STEPS site coordinator and / or the STEPS coordination committee.	
11	Provide feedback.	

**Note:** The main tasks are further described below.

---

**Train staff** Train the data collection team staff in:

- interview skills
- approaching households
- keeping track of interviews and non-response
- informing participants and obtaining consent
- conducting interviews for Step 1
- taking measurements for Step 2
- completing the instruments
- giving feedback to the participant
- using the forms and available tools, including making appointments for Step 3.

**Note:** Details on training are discussed in Part 3, Sections 2-4.

---

*Continued on next page*

## Supervising Data Collection, Continued

---

### **Assign interview teams to locations**

Create a list of the areas to be surveyed and assign data collection teams to each location. When assigning locations:

- schedule interview teams to survey one location before moving to another;
  - schedule time to revisit each location to finish interviews;
  - keep a record of all interviewers that need transport and schedule the transport;
  - keep track of which locations were visited by which interview team.
- 

### **Contact local authorities**

The data collection team supervisor will need to contact appropriate local authorities to inform them about the survey and gain their support and cooperation.

---

### **Obtain and prepare data collection forms and questionnaires**

Ensure there are sufficient quantities of the printed instruments, Kish Household Coversheets, Interview Tracking Forms, and all necessary forms and STEPS tools required for the interviewers to use. Please see Part 6, Section 2 for all available data collection templates.

Prior to beginning data collection, the Kish Household Coversheet as well as the Interview Tracking Form must have specific parts completed. 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. See the final topic of Part 2, Section 2, entitled "Preparing Data Collection Forms", for specific instructions on how to partly fill out the collection forms to prepare field work.

In addition, the following parts of the instruments should be completed prior to beginning data collection:

- Participant ID on each page
  - Cluster ID
  - Cluster name
  - Interviewer ID.
- 

### **Distribute Data Collection Forms**

Distribute to each interviewer all the instruments, data collection forms and equipment required prior to interviewers going into the field.

---

### **Tracking Interviews**

Interviewers should use the Interview Tracking Form, available in Part 6, Section 2, to track household and participant response information on a daily basis. Collect the forms on a regular basis and give to the data entry team supervisor.

For other details not included on these forms, set up a log book.

---

*Continued on next page*

## Supervising Data Collection, Continued

---

**Supervise data collection** To ensure high-quality data collection, the supervisor will need to observe a certain proportion of the interviews conducted by each interviewer, particularly at the beginning of the data collection period.

The proportion may vary depending on the interviewers' experience, the timeframe and the budget involved.

The supervisors should also check that each instrument has been completed properly.

Ensure all instruments and other forms are accounted for and in order before sending them to the STEPS office for data entry.

---

**Manage Human Resources** Manage and support the data collection team to ensure :

- good quality interviews are conducted and Instruments are complete
  - interview timeframes are adhered to
  - interviewers are supported if participant issues arise
  - performance issues are dealt with appropriately
  - confidentiality of all STEPS surveillance material is respected at all times
  - feedback is provided to data collection staff
  - sick leave and annual leave is appropriately covered.
- 

**Progress reports** During the data collection stage, you will need to provide regular updates to the STEPS site coordinator and / or the coordination committee. This should include:

- updates on progress against scheduled data collection timeframes
  - issues and problems encountered.
- 

**Feedback** When data collection is completed, get together with the data collection teams to debrief and gain valuable feedback. This will be useful for processing and analysing the data and for revising the Instrument and manuals for the next round of STEPS surveillance.

---

## Data Collection Process and Interviewer Tasks

---

### Introduction

Data collection starts in the field only when the actual planning of the STEPS survey has been done. Each of the stages for data collection needs to be undertaken appropriately to ensure accurate data is being collected.

Interviewers have a key role to play in the STEPS surveillance process. The quality of the data collected and therefore the available results depends on successful interviews.

---

### Interviewer Tasks during Data Collection Process

An overview of the tasks of an interviewer are included in the following checklist.

Task	Description	✓
1	Door knock selected households.	
2	Brief household members on purpose of the survey.	
3	Record all eligible participants on the Kish Household Coversheet and select one using the Kish Method.	
4	Record information on the Interview Tracking Form.	
5	Inform the selected participant using the Participant Information Form and obtain written consent.	
6	Conduct the interview and record results for Step 1.	
7	Double check completed Step 1 questions.	
8	Take measurements and record results for Step 2 (if applicable).	
9	Double check completed Step 2 information.	
10	Fill in Participant Feedback Form on results of Step 2 measurements for the participant.	
11	Make appointment for Step 3 (if applicable) and inform participant on fasting.	
12	Check all completed forms and hand to supervisor.	
13	Report any difficulties to supervisor.	

**Note:** Each of these tasks is described in more detail below and in Part 3, Sections 2-4.

---

*Continued on next page*

## Data Collection Process and Interviewer Tasks, Continued

**What the interviewer will need**

The forms and resources the interviewer will need for data collection are listed in the following checklist:

For Step			Form	✓
1	2		Map or list of households in sample	
1	2		Name tag	
1	2		Notification of WHO STEPS surveillance visit	
1	2		Kish Household Coversheet	
1	2	3	Participant Information Form	
1	2		Consent Form 1	
		3	Consent Form 2	
1	2		Interview Tracking Form	
1	2	3	STEPS Instrument	
1	2	3	Question-by-Question Guide	
1	2		Show cards	
		3	Clinic Appointment Card (with map if necessary)	
		3	Fasting instructions	
	2		Participant Feedback Form (Step 2)	
		3	Participant Feedback Form (Step 3)	
		3	Clinic Registration Form	

## Approaching Selected Households and Participants

**Introduction** For Step 1 and Step 2 of the instrument, the interviewers will need to physically visit individual households to conduct the survey.

**Contact process** See the table below for an overview of the contact process.

Stage	Description						
1	Obtain household lists with associated addresses (and map if necessary) from your supervisor.						
2	Physically knock on the door.						
3	<table border="1"> <thead> <tr> <th>If ...</th> <th>Then ...</th> </tr> </thead> <tbody> <tr> <td>Nobody is home</td> <td>Leave a Notification Card and record on the Interview Tracking Form.</td> </tr> <tr> <td>Somebody is home</td> <td>Introduce yourself and exchange greetings.</td> </tr> </tbody> </table>	If ...	Then ...	Nobody is home	Leave a Notification Card and record on the Interview Tracking Form.	Somebody is home	Introduce yourself and exchange greetings.
If ...	Then ...						
Nobody is home	Leave a Notification Card and record on the Interview Tracking Form.						
Somebody is home	Introduce yourself and exchange greetings.						
4	Explain the reason for your visit and purpose of the STEPS surveillance.						
5	Record each person living in the house between the ages of 25-64 on the Kish Household Coversheet.						
6	Select household participant using the Kish Household Coversheet (unless pre-selected).						

**Note:** Each of these stages is described in more detail below.

**Door knocking procedure** Contact attempts must be made by actually knocking on the door of the household, simply walking by and thinking that no one is at home cannot be counted as an attempted contact.

Use the following table to help with different situations when you knock on the door.

If...	Then...
Someone is at home	Speak to the first adult you encounter in the household.
Nobody answers the door-knock	Look round side of house to see if someone is nearby.
Nobody is at home	Leave a notification of WHO STEPS surveillance visit and record details in the Interview Tracking Form (see below how to complete this form).
Household members are not available at the time of the first visit.	Make at least 2 different visits to obtain an interview. Choose times that are different – early morning or late afternoon.

*Continued on next page*

## Approaching Selected Households and Participants, Continued

---

**Recording household details** Record if anyone is home and the date and time of the visit on the Interview Tracking Form. See "Completing the Interview Tracking Form" on page 4-1-12 below.

---

**Introducing yourself** Make sure your name tag is attached and clearly visible.  
Introduce yourself and explain the reason for your visit as follows:

My name is \_\_\_\_\_ and this is \_\_\_\_\_. We are employees of the <Ministry of Health> and we are working in a team to conduct a survey on health issues. We are hoping that the people in this house will participate in this survey. We would like to find out the number of people usually residing in this house between the ages of 25-64. Can you please give me the first name of those who usually live in this house between the ages 25-64 (starting, for example, with the oldest male)?

---

**Explaining purpose of the survey** Explain that the purpose of this study is to determine the extent of chronic noncommunicable diseases (i.e. long-standing diseases not caused by infections) major risk factors in your country. These risk factors include:

- tobacco use
- alcohol consumption
- low intake of fruit and vegetable
- physical inactivity
- obesity
- raised blood pressure
- raised fasting blood glucose
- high levels of fat in the blood.

Explain that once the study data has been collected and analysed, this will help your health services plan and determine public health priorities to:

- prevent chronic disease epidemics before they occur
  - monitor and evaluate population-wide chronic disease programmes.
- 

**Record all eligible household members** For each eligible person in the household record the following details on the Kish Household Coversheet (see Part 6, Section 2):

- sex
  - age.
- 

**Select participant using Kish Method** Select one participant from the household using the Kish Household Coversheet (see Part 6, Section 2). Details on how to select a participant in a household are provided in Part 2, Section 2, under "The Kish Method".

---

## Informing Participants

---

**Introduction** After having chosen a study participant from a household, this participant needs to be informed on the details of the study before he/she will be asked to sign the Consent Form. For informing the participant, the Participant Information Form can be read out (see Part 6, Section 2).

---

**Explaining aim of the survey** Explain that the aim of the survey is to determine population levels of major chronic disease risk factors. Also explain how the information will be used, i.e. for policy making in order to decrease risk factor levels.

---

**Explaining survey process** Explain that you will collect information from a number of pre-selected households throughout the country. Explain how data will be collected, as appropriate, i.e. through:

- interview questions (Step 1)
- measurements of height, weight, waist, and blood pressure (Step 2)
- blood tests for sugar and fats (Step 3).

---

**Explaining collection methods** Use the table below to help run through the whole data collection process with the participant:

Stage	Description
1	Step 1, asking questions about participant's: <ul style="list-style-type: none"><li>• age;</li><li>• education;</li><li>• employment (if appropriate);</li><li>• income (if appropriate);</li><li>• tobacco and alcohol use;</li><li>• fruit and vegetable intake;</li><li>• physical activity;</li><li>• knowledge and history of high blood pressure and diabetes (if appropriate).</li></ul>
2	Step 2, taking the following measurements: <ul style="list-style-type: none"><li>• height and weight</li><li>• waist circumference</li><li>• blood pressure</li><li>• hip circumference (if appropriate)</li><li>• heart rate (if appropriate).</li></ul>
3	Step 3 (if appropriate), taking a small amount of blood from a prick on your finger or a vein in your arm to determine blood sugar and blood lipid levels.*  * <b>Note:</b> This may cause some mild pain
4	Respond to any questions the participant may have.

---

*Continued on next page*

## Informing Participants, Continued

---

### Survey timeframe

It is estimated that each part (i.e. Step 1, Step 2 and then, Step 3) of the survey will take approximately the following timeframes:

Step	Timeframe
1	30 minutes
2	30 to 45 minutes
3	5 minutes

---

### Other items to explain to participants

Use the table below to help explain to each participant the benefits, their rights and how confidentiality will be handled.

In terms of...	You will need to explain to each participant that...
Community benefits	The results of this study will be used to assist the Ministry of Health develop public health programs that target efforts to lower the risk factors that lead to chronic diseases.
Individual rights	Participants may: <ul style="list-style-type: none"><li>• decline to take part in the study;</li><li>• withdraw their consent at any time;</li><li>• not answer any questions in the interview that they do not wish to answer.</li></ul>
Confidentiality	<ul style="list-style-type: none"><li>• Participants should provide their name and contact information so they can be contacted if there is any problem following the analysis of the information and follow-up is necessary.</li><li>• Participation and data provided will be completely confidential.</li><li>• While the data from this study may be sent elsewhere for analysis, no personally identifiable information will be provided for this analysis.</li><li>• Their name and their household or village will not be used in any report of the study.</li></ul>

---

## Obtaining Consent

**Introduction** Each participant must provide both verbal and written consent before taking part in the survey.

**Obtain consent** For those who will take part in the study, follow the steps below to obtain verbal and written consent.

Step	Action								
1	<p>Use the following table to select the appropriate consent form for each person taking part:</p> <table border="1"> <thead> <tr> <th>In...</th> <th>Then use Consent Form...</th> </tr> </thead> <tbody> <tr> <td>Step 1 only</td> <td>1</td> </tr> <tr> <td>Step 1 and 2 only</td> <td>1</td> </tr> <tr> <td>Step 1, 2 and 3</td> <td>1 and 2</td> </tr> </tbody> </table> <p><b>Note:</b> See Part 6, Section 2 for suggested drafts of consent forms.</p>	In...	Then use Consent Form...	Step 1 only	1	Step 1 and 2 only	1	Step 1, 2 and 3	1 and 2
In...	Then use Consent Form...								
Step 1 only	1								
Step 1 and 2 only	1								
Step 1, 2 and 3	1 and 2								
2	<p>For each participant, use two copies of the consent form(s) as follows:</p> <ul style="list-style-type: none"> <li>• one for the participant to keep</li> <li>• one for the STEPS coordination office.</li> </ul>								
3	<p>Allow the participant to read the consent form(s) or, in case of poor eyesight or illiteracy, read it out to them.</p>								
4	<p>Use the table below to help with the following situations:</p> <table border="1"> <thead> <tr> <th>If...</th> <th>Then...</th> </tr> </thead> <tbody> <tr> <td>The intended participant answers NO to any question in the consent form.</td> <td>Ask the participant whether he/she understands the questions.</td> </tr> <tr> <td>The participant does not understand the question.</td> <td>Rephrase the question.</td> </tr> <tr> <td>The participant understands the question and the answer is still NO.</td> <td>Circle NO in the consent form and record age and sex as best you can.*</td> </tr> </tbody> </table> <p>*This means that the household member will not participate in the survey. However, you must still include him / her in the Interview Tracking Form, then move to the next selected household.</p>	If...	Then...	The intended participant answers NO to any question in the consent form.	Ask the participant whether he/she understands the questions.	The participant does not understand the question.	Rephrase the question.	The participant understands the question and the answer is still NO.	Circle NO in the consent form and record age and sex as best you can.*
If...	Then...								
The intended participant answers NO to any question in the consent form.	Ask the participant whether he/she understands the questions.								
The participant does not understand the question.	Rephrase the question.								
The participant understands the question and the answer is still NO.	Circle NO in the consent form and record age and sex as best you can.*								
5	Get the participant to sign both copies.								
6	As the interviewer, you must sign as a witness.								
7	Thank him/her for agreeing to take part in the survey.								

## Completing the Interview Tracking Form

---

**Introduction** You need to record every household visited on the Interview Tracking Form.

For a copy of the Interview Tracking Form, see Part 6, Section 2.

---

**Purpose of Interview Tracking Form** The purpose of the Interview Tracking Form is to document and be able to report on:

- number of households visited;
- number of eligible individuals in each household;
- Participant ID;
- if the participant was at home on either the first or second visit;
- age group and sex of the participant;
- participant eligibility for Step 1, Step 2, and Step 3 and if they consented or declined each step;
- appointment date and time for a scheduled interview (in case participant was not at home at the first visit);
- individual comments.

**Note:** The Interview Tracking Form is used during analysis. If this form is not used, you will not be able to properly weight your data which will reduce the quality of your results.

---

**Completion guidelines** Depending on the sample design the Interview Tracking Form may already be partially completed (see Part 2, Section 2 "Preparing Data Collection Forms"). Use the following table for guidance on how to finish completing this form.

Column	Guidelines for completion
Cluster ID	ID code associated with the cluster. Separate forms need to be used for different clusters.
Household ID	Use the predetermined codes, see Part 2 Section 2.
Number eligible in household	Record the number of eligible people (aged 25 to 64) in the household.
Participant ID	Mark the Participant ID that is written on the instrument to be used.
At home (visit 1 and visit 2)	<ul style="list-style-type: none"><li>• If participant is at home, then mark "Y".</li><li>• If participant is not at home, then mark "N".</li></ul>
Male/female by age group	Mark an "X" in the box according to the sex and age group of the participant.

*Continued on next page*

## Completing the Interview Tracking Form, Continued

---

### Completion guidelines (cont.)

Column	Guidelines for completion
Step 1 (Yes, Decline)	<ul style="list-style-type: none"> <li>• Mark an "X" if participant consents to the interview (yes column).</li> <li>• Mark an "X" if participant declines.</li> </ul>
Step 2 (Yes, Decline)	<ul style="list-style-type: none"> <li>• Mark an "X" if participant consents to the interview (yes column).</li> <li>• Mark an "X" if participant declines.</li> </ul>
Step 3 (Yes, Decline)	<ul style="list-style-type: none"> <li>• Mark an "X" if participant consents to the interview (yes column).</li> <li>• Mark an "X" if participant declines.</li> </ul>
Appointment Time	If you schedule an appointment with a participant, record the date and time here.
Individual Comment	<p>Free area for interviewers to record comments. Some reasons to use this field may be that the participant:</p> <ul style="list-style-type: none"> <li>• has a communication problem (e.g. speaks a local dialect only, has hearing impairment);</li> <li>• refuses to consider participation;</li> <li>• is ill, cannot obtain consent;</li> <li>• has a disability;</li> <li>• cannot miss work;</li> <li>• refuses to take part in Step 3 (e.g. is afraid of needles or has cultural/religious preference not to provide blood).</li> </ul>

### Notes:

- If you altered the age range of your survey you will need to reflect those changes on the interview tracking form (for example, if you sample 15-24 year olds you will need to add 2 other columns to the tracking form).
  - If your country/site does not do Step 3 measurements, the Interview Tracking Form should be altered by removing the according columns.
  - If your country/site does Step 3 with a sub-sample, a column "not applicable" should be added under Step 3.
-

## Recording Information

---

**Introduction** All results that are recorded on the instruments must be written as clearly as possible to avoid ambiguity and confusion when checking and entering the results.

---

**Requirements** Some general requirements for recording survey information are as follows:

- use a pencil, not a pen for writing;
  - record the identification number on each instrument and each page;
  - do not erase any notes made;
  - if a question has been skipped by mistake, correct it;
  - if a participant changes his / her mind on one of the options, record the new answer;
  - record only answers that are relevant to the survey;
  - record comments or explanations in brackets in the instrument next to the corresponding question;
  - don't get too absorbed recording, keep the participant's interest by saying the participant's response aloud as you write it down;
  - reach an agreement on how to write numbers (notably 1s and 7s).
- 

**Handling issues** Use the table below to help with some common issues you may encounter.

<b>If...</b>	<b>Then...</b>
You are uncertain about a response.	Repeat the question and record the answer exactly. Do not paraphrase a response.
If the participant doesn't know the answer to a question.	Record 77 or 777 (depending on number of spaces) for "don't know".
You have missed a question.	Go back and ask the question, making a note in the margin that the question was asked out of sequence.
Missing data is not discovered until after the interview.	If possible, re-contact the participant and ask the question. Note in the margin that the question was asked out of sequence. If not possible, for example the team has moved on from the village, then record 99 or 999 (depending on the number of spaces) and make a note that the question was skipped due to interviewer error.
The participant refuses to answer a question.	Record 88 or 888 (depending on the number of spaces). <b>Note:</b> Before accepting a refusal explain the objective of the question to the participant.

---

*Continued on next page*

## Recording Information, Continued

---

### **Checking and editing**

Before leaving the household or the clinic setting, check the instrument and make sure that:

- all the questions have been answered and all the measurements have been taken (if applicable);
  - the information recorded is clear and legible;
  - probing comments are indicated;
  - all the information has been completed including the Participant ID on every page.
- 

### **Interviewer's Guide**

See more information on how to record information in Part 3, Section 2 "Completing the STEPS Instrument".

---

## Completing the Participant Feedback Form

---

### **Introduction**

After having completed the Step 2 measurements, the participant should be informed on his/her results. You can use the Participant Feedback Form (Step 2) in order to give the participant feedback on his body measurements (see Part 6, Section 2). This form stays with the participant after having completed the survey.

---

### **Filling in the Participant Feedback Form**

Please follow the following guidelines when completing the Participant Feedback Form:

- blood pressure: record the average of reading 2 and 3 for both systolic and diastolic blood pressure  $((\text{reading } 2 + \text{reading } 3) / 2)$ ;
  - blood pressure classification: tick the appropriate box;
  - heart rate: record the average beats per minute of reading 2 and 3  $((\text{reading } 2 + \text{reading } 3) / 2)$ ;
  - height and weight: record height in cm and weight in kg;
  - body mass index: calculate the body mass index and record (weight in kg divided by meters squared:  $\text{kg}/\text{m}^2$ ), the BMI Classification Chart helps calculating the BMI (see Part 6 Section 2);
  - BMI classification: tick the appropriate box, the BMI Classification Chart helps finding the BMI category;
  - waist and hip circumference: record waist and hip circumference (if applicable) in cm.
-

## Scheduling Clinic Visits for Step 3 Measurements

---

### Scheduling Step 3 measurements

If your site plans to take biochemical measurements for Step 3, you will need to schedule those for participants that have been selected to visit the clinic for tests.

---

### Scheduling

Follow the steps below to schedule and brief participants:

Step	Action
1	Ask the participant the day and time they would like to come in to the clinic or designated place for blood tests (Step 3) using the times assigned to your team.
2	If necessary, provide a map showing the venue.
3	Record the time in the appropriate box on the Clinic Appointment Card (see Part 6, Section 2). Leave this card with the participant. Take note of the appointment time for communication to the data collection team supervisor.
4	Provide a copy of the Fasting Instructions and explain the importance of fasting properly.
5	Remind the participant to bring to the clinic their own copy of the signed consent form as well as the Clinic Appointment Card as a means of identification.
6	In cases where participants need transportation to the clinic or designated place for blood tests, make the arrangement and inform your supervisor.

---



## Section 2: Data Entry

### Overview

---

**Introduction** This section covers all the tasks that need to be conducted to enter the STEPS surveillance data as recorded on the STEPS Instrument and check and correct data errors.

---

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

- data entry supervisor
  - data entry staff
  - STEPS Site Coordinator
  - data analyst.
- 

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

Task Name	Duration	Month 3	Month 4	Month 5
Enter data (1 <sup>st</sup> and 2 <sup>nd</sup> key entry)	8 weeks			
Check and clean data	8 weeks			
Merge data	1 day			

---

**In this section** This section covers the following topics:

Topic	See Page
Supervising Data Entry	4-2-2
Data Entry	4-2-6
Backup and Filing	4-2-8
Reporting	4-2-9
Creating the Final Dataset	4-2-10

---

## Supervising Data Entry

---

**Introduction** Members of the data entry team may have different levels of skills, experience and varying strengths and abilities. To ensure high standards in this environment, the appointment of one person to lead the team and supervise the work is necessary.

---

**Core tasks** The core tasks of the data entry supervisor are listed in the table below. General roles are identified in Part 1, Section 2.

Tasks	Description
1	Train data entry staff in daily operations.
2	Receive, log and assign instruments from data collection team.
3	Assign data entry staff to computers and instrument components for data entry.
4	Create folders with coversheets to track and manage entry of instrument data.
5	Distribute and manage folders on a daily basis.
6	Check and correct data entry anomalies and review consistency reports.
7	Regularly liaise with data collection team supervisors.
8	Supervise human resources and hold regular team meetings.
9	Supervise technical requirements including: <ul style="list-style-type: none"><li>• daily backup of data on each computer</li><li>• regular virus scanning and virus software updates</li><li>• software support.</li></ul>
10	Provide regular progress reports to the STEPS site coordinator and/or the STEPS Coordination Committee.
11	Create final data set.

**Note:** Tasks 1-8 are further described below. Task 11 is described on page 4-2-10.

---

**Training** For more information on training data entry staff, please see Part 3, Sections 1 and 5.

---

*Continued on next page*

## Supervising Data Entry, Continued

---

### Receive, log, and assign Instruments

At the end of each day, data collection team supervisors should supply all completed instruments to the data entry office. The data entry supervisor should receive and log them as follows:

Step	Action
1	Log receipt of instruments in the data_entry_log.xls file.
2	Assign each instrument to a specific computer and record this assignment in the data_entry_log.xls file.

---

### Assign computers and staff

Set up the data entry computers and assign staff so that:

- the Interview Tracking Form information is entered into the interview\_tracking\_form.xls files;
  - the main instrument data (i.e. Step 1, Step 2 and Step 3) and interview location and date information are entered into survey.rec and biochemical.rec (if applicable);
  - identifying information is entered into consent.rec (if applicable);
  - different data entry staff are assigned to conduct first key and second key entries on each computer (staff can swap computers to do the second keying).
- 

### Create folders (or boxes)

For each data entry computer, prepare three folders (or boxes), each with a coversheet (see coversheet template Part 6, Section 2) for the different stages of data entry as follows:

Folder	For Instrument data that...	Folder name
1	Is not yet entered.	1 <sup>st</sup> key
2	Has first keying complete.	2 <sup>nd</sup> key
3	Has second keying complete.	Completed

**Note:** Once folders have been assigned to a computer for data entry, they should remain with that computer.

---

*Continued on next page*

## Supervising Data Entry, Continued

---

### Distribute and manage folders

Follow the steps below to distribute and manage the folders on a daily basis.

Step	Action
1	Each morning give data entry staff their computer folders and a pile of instruments to enter from your general folders/boxes.  <b>Note:</b> We suggest 30-40 instruments per staff or as many as can be realistically entered by one person in one day.
2	Log distribution of new instruments in the data_entry_log.xls file.
3	At the end of each day, collect all folders from each data entry staff, check that the coversheet is attached and labeled, and lock these away.

---

### Check and correct anomalies

Check and handle all data entry errors and corrections by:

- making decisions on alterations to data from completed instruments where necessary;
  - ensuring data entry staff use the Data Entry Tracking Form (located in Part 6, Section 2) to record all ambiguous or unclear data entry results, questions and problems;
  - clearly annotating original forms for an audit trail;
  - referring to data collection staff where necessary;
  - creating a list of potential problems and frequently asked questions (FAQs);
  - working with data analyst (where appropriate) to systematically work through data anomalies.
- 

### Consistency report

Review weekly consistency reports for each data entry computer to detect problematic data. Data entry staff should be trained how to produce a consistency report for the data entered on their computer. Instructions for producing a consistency report are provided on Part 3, Section 5.

---

### Liaise with data collection team

Once data entry has started, you should have regular discussions with the data collection team supervisor to provide feedback on:

- data quality
- workflow and receipt of instruments
- issues and anomalies that may arise.

See page 4-2-9 for further details on reporting back to the data collection team supervisor.

---

*Continued on next page*

## Supervising Data Entry, Continued

---

### **Manage human resources**

Manage and support the data entry team to ensure :

- good workflow
  - high quality of data entry
  - high level of cooperation between team members
  - different data entry operators are scheduled for first and second key entry
  - performance issues are dealt with appropriately
  - confidentiality of all STEPS surveillance material is respected at all times
  - feedback is provided to data entry staff
  - sick leave and annual leave is appropriately covered.
- 

### **Team meetings**

Schedule weekly meetings to discuss the data entry process and report results to the data entry team. These meetings should be used to:

- discuss problems or concerns
  - collect and review data entry tracking forms
  - collect and review consistency reports
  - discuss progress in data entry process.
-

# Data Entry

---

**Introduction** STEPS survey data on the completed instruments is to be entered by trained data entry staff into a series of predefined data entry templates using EpiData.

---

**EpiData templates** For Step 1, 2 and 3 you will need to use the following data entry templates:

- Consent (if applicable)
- Survey
- Biochemical\*

\* If Step 3 results are recorded on forms separate from the main STEPS Instrument, then the biochemical template will need to be used. Otherwise, biochemical results are entered into the Survey template.

---

**Interview Tracking Form Excel File** Data from the Interview Tracking Forms are to be entered into the interview\_tracking\_form.xls Excel file. See Part 2, Section 4 for downloading information.

It is recommended that all Interview Tracking Forms be entered on the Master Computer by the data entry supervisor.

**Note:** This Excel spreadsheet should not be modified because it will be imported into your database and used to calculate the non-response weight for your data. If any modifications are needed (e.g. additional age groups or fewer age groups) please contact the STEPS team for assistance.

---

**Data entry process** Data entry is a systematic process that covers the following main stages:

Stage	Description
1	Receiving, logging and assigning
2	First key data entry
3	Second key data entry
4	Validation and error correction
5	Backing up
6	Storing and filing the instruments

---

*Continued on next page*

## Data Entry, Continued

---

### Receiving, logging and assigning

The data entry supervisor should log all received instruments on a daily basis and distribute these to data entry staff. See page 4-2-3 for further details.

Data entry staff are responsible for regularly updating the Data Entry Tracking Form for their computer, ensuring that for each instrument the current status is correct and any data entry problems are recorded.

---

### Entering Data

At each computer, the data entry process should proceed as follows:

Step	Action
1	Complete first keying on all instruments in "1 <sup>st</sup> key" folder and move each entered instrument to "2 <sup>nd</sup> key" folder.
2	After all instruments in "1 <sup>st</sup> key" folder have been entered, prepare for second keying.
3	Complete second keying and move each entered instrument to "Completed" folder.

Specific instructions for each of these steps is covered in detail in Part 3, Section 5. Prior to beginning data entry, it is essential that data entry staff have been thoroughly trained and the procedures for handling potential problems during the data entry process have been well established.

---

### Validation and error correction

During second keying, if there are any discrepancies between the data from the first keying and data from the second keying the data entry program will immediately inform you that the data does not match. Part 3, Section 5 provides specific instructions for handling these discrepancies. Prior to beginning data entry, data entry staff must be thoroughly trained to know exactly what to do when these discrepancies arise.

---

### Backing up data

Every computer should be backed up at the end of each day. It is the responsibility of the data entry team member to back up the computer they used. For more detailed information on backing up the computers see Part 3, Section 5.

---

### Storing and filing the Instruments

At the end of each day all folders should be placed in a secure location. More detailed information on filing can be found on the following page.

---

## Backup and Filing

---

**Introduction** All files associated with the STEPS survey need to be properly filed and all electronic data backed up on a daily basis to avoid data loss.

---

**Filing completed STEPS Instruments** At the end of each day of data entry, all the instruments that have been entered need to be filed in the appropriate folders (first keying, second keying, or completed) designated for each computer.

All data that has not been entered needs to be returned to the supervisor to be stored in the collective folders.

---

**Backup** At the end of each day of data entry, each data entry staff member should backup all data files on their computers. This is to avoid data loss. Detailed instructions for backing up the data files using EpiData are provided in Part 3, Section 5.

---

**Archiving** For details on archiving all STEPS surveillance files, please see Part 4, Section 5.

---

## Reporting

---

**Introduction** The data entry supervisor should regularly liaise with and report progress and issues to the:

- data collection team supervisor
  - STEPS Site Coordinator
  - STEPS Coordinating Committee.
- 

**What to report to whom** Use the table below for guidance on what to report to whom.

<b>What to report</b>	<b>To whom</b>	<b>When</b>
<ul style="list-style-type: none"><li>• errors on completed instruments</li><li>• data collection timeframes not being met</li></ul>	Data collection team supervisors	At least weekly
<ul style="list-style-type: none"><li>• progress</li><li>• issues that need resolving</li><li>• timeframe and budget updates</li></ul>	STEPS Site Coordinator	Weekly
<ul style="list-style-type: none"><li>• progress</li><li>• timeframe and budget updates</li></ul>	STEPS Coordinating Committee	Monthly

---

## Creating the Final Dataset

---

**Introduction** Once all the survey data have been entered twice and checked on each computer, all the files used on each computer need to be combined into a single dataset (except the consent information) so the data can be analysed.

In EpiData this process is called appending and merging.

---

**Requirements** This process should be conducted:

- by the supervisor (or a single senior staff member)
  - on the Master computer
  - after all data has been backed up.
- 

**Process** The appending and merging process includes the following stages:

Stage	Description
1	Listing each data entry computer and what data was entered on it
2	Performing a record count for each .rec file on each data entry computer
3	Copying files to the Master computer
4	Appending the data
5	Merging the datasets
6	Checking the final dataset

---

**Create list of computers** In a log book or somewhere safe, list each data entry template and the computers on which it was entered. For ease of use, list the computers in alphabetical order starting with Master, then A, B, C, etc., for each template. For example:

Template	Computer
Survey	<ul style="list-style-type: none"><li>• Master</li><li>• A</li><li>• B</li><li>• C</li></ul>
Biochemical	<ul style="list-style-type: none"><li>• D</li><li>• E</li></ul>

---

*Continued on next page*

## Creating the Final Dataset, Continued

### Record count on each computer

Perform a record count for each .rec file on each data entry machine to check the record count. Follow the steps below to perform a record count:

Step	Action
1	Open EpiData.
2	Select "Count Record" from the "Document" menu at the top of the screen.
3	Select one .rec file of interest and click 'add to list'.
4	Repeat step 3 until all the .rec files on the computer have been added to the list.
5	Select 'id' in the 'field to evaluate' list.
6	Select "OK".
7	The screen will then show a list of all IDs used in any of the .rec files selected and indicate the number of instances of each ID in each .rec file selected.
7	Click Save and save document under C:/STEPS/data with the title recordcount + machine label (e.g. recordcountA).

### Copying files to the Master computer

All the data files that have been saved under the STEPS folders on each computer must be copied to the Master computer so they can be appended and merged. Follow the steps below to copy the files.

Step	Action
1	Create a new folder on the Master computer labeled: MainSTEPS (C:\MainSTEPS).
2	Under C:\MainSTEPS create a folder for each data entry computer as follows: <ul style="list-style-type: none"><li>• C:\MainSTEPS\Master</li><li>• C:\MainSTEPS\A</li><li>• C:\MainSTEPS\B, etc.</li></ul>
3	Make a copy of the C:\STEPS\data folder (on the Master computer) and add to the C:\MainSTEPS\Master folder.
4	Copy and paste the C:\STEPS\data folders from each machine (one at a time) and place in the folder that is labeled with the original machine name (e.g. the C:\STEPS\data folder from computer A would be copied to C:\MainSTEPS\A on the Master computer).

**Note:** Never move the folders to the Master computer, create copies only. This will allow you to backtrack in case of errors.

*Continued on next page*

## Creating the Final Dataset, Continued

### Appending the data

All the files stored on the Master computer must be appended to combine the data from each computer (A, B, C, etc.) into single files. Follow the steps below to append your data. At the end of this process you should have one combined data (.rec) file for each template used.

Step	Action
1	In EpiData, select "Append/Merge" from the "Data in/out" menu at the top of the screen.
2	In the dialog box, click "Name of first data file" and select the Master computer file. If the data entry template was not entered on the Master computer, select the first computer in your list on which the template was used.
3	Click on "Select name of second data file" and select the next computer listed in your list of computers on which the template was used.
4	Click "OK".
5	Type the merged data file name (e.g. SurveyMasterA) in the "Resulting data file" field.
6	Click on "append only data fields in data file B that also exist in data file A" on the "Append" tab in the lower left-hand corner of box.
7	Click "Append" and enter data file description (e.g. Appended Master + A).
8	Make note of the second paragraph of the Information window "Appended and saved as:". This tells you the name and location of the appended file.
9	Repeat steps 3-8, using the new file created each time. Repeat the steps until each computer's files have been appended into one master file (e.g. SurveyMasterABCDEF).
10	Upon completing these steps for all data entry templates used, you should have one .rec file for each template used in your C:\STEPS\MainSTEPS folder.  Rename each of these .rec files "Master" followed by the name of the specific template. Your C:\STEPS\MainSTEPS folder should now have: <ul style="list-style-type: none"><li>• MasterConsent.rec (if used)</li><li>• MasterSurvey.rec</li><li>• MasterBiochemical.rec (if applicable)</li></ul>

*Continued on next page*

## Creating the Final Dataset, Continued

### Checking the appended data

Check the results of the append process for each master data file. Each master data file should have the same number of records, which should equal the total number of correctly completed instruments.

Follow the steps below to check the appended data files:

Step	Action
1	Perform a record count for each appended master data file you have just created (see page 4-2-11 for instructions on performing a record count).
2	Using the record counts obtained earlier from each computer, sum the record counts across all computers for each template.
3	Confirm that the record count is the same number for each master data file and that this number equals the sum calculated from all data entry computers. If there are any discrepancies, further investigation is needed before continuing with the creation of the final data set.

### Merging files into one dataset

The table below describes how to merge the master data files containing the appended data into one master dataset. This will only need to be done if you used both the survey.rec and biochemical.rec templates.

Step	Action
1	In EpiData, select "Append/Merge" from the "Data in/out" menu at the top of the screen.
2	In the dialog box, click "Name of first data file" and choose C:\STEPS\MainSTEPS\MasterSurvey.rec.
3	Click "Select name of the second data file" and choose C:\STEPS\MainSTEPS\MasterBiochemical.rec.
4	Click "OK".
5	Type the merged data file name "MasterDataSet" in the "Resulting data file" field.
6	Click on the "Merge" tab in the lower left-hand corner of box.
7	Select "Merge only records from data file B that match records in data file A".
8	Select ID in the "Select key field(s)" list.
9	Click "Merge" and enter data file description (e.g. Merged survey + biochemical).
10	Make note of the second paragraph of the Information window "Merged and saved as:". This tells you the name and location of the merged file.

**Note:** If the consent information was entered electronically (which is not recommended) **DO NOT** merge this information into the MasterDataSet.

*Continued on next page*

## Creating the Final Dataset, Continued

### Checking the merged data

Check the results of the merge process by evaluating the record count and making sure all variables are included. Follow the steps below to check your MasterDataSet.rec file.

Step	Action
1	Perform a record count for MasterDataSet.rec (see page 4-2-11).
2	Compare this figure to the sum of the record counts obtained from all data entry machines.
3	If results do not match, further investigation is needed.
4	Open the MasterDataSet.rec file in EpiData by selecting "Open" from the "File" menu at the top of the screen.
5	Scroll down through the data entry screen to confirm that fields from all used templates are present in this file, if any fields are missing, further investigation is needed.

### Export the dataset for analysis

The dataset needs to be exported into a format that is readable by Epi Info. Follow the steps below to export the dataset.

Step	Action								
1	Open EpiData.								
2	Click on "6. Export Data".								
3	Select "dBase III" from the list.								
4	Select MasterDataSet.rec as the file to open.								
5	In the "Export data file to dBase III file" window: <table border="1"><thead><tr><th>Step</th><th>Action</th></tr></thead><tbody><tr><td>5.1</td><td>Select "all records".</td></tr><tr><td>5.2</td><td>Check "skip deleted records".</td></tr><tr><td>5.3</td><td>Click "All" from the Select Fields section.</td></tr></tbody></table>	Step	Action	5.1	Select "all records".	5.2	Check "skip deleted records".	5.3	Click "All" from the Select Fields section.
Step	Action								
5.1	Select "all records".								
5.2	Check "skip deleted records".								
5.3	Click "All" from the Select Fields section.								
6	Click "OK".								

**Note:** There are many different export formats available from EpiData. If you are using another analysis package other than Epi Info you will need to select the appropriate format from the list of formats offered in step 3.

*Continued on next page*

## Creating the Final Dataset, Continued

---

### **Preparing the interview tracking form Excel file**

The interview\_tracking\_form.xls file should be on the Master computer and it should contain the information from all completed Interview Tracking Forms. If the data was entered on multiple computers then you will need to merge all the different documents into one spreadsheet by copying and pasting the records from one spreadsheet into another.

Once you have confirmed that all interview tracking information is in one interview\_tracking\_form.xls file on the Master computer, click on the "Format for Epi Info" button in the "Instructions" worksheet of the file to generate the worksheet "EpiInfo" to be used during analysis.

The "EpiInfo" in the interview\_tracking\_form.xls file worksheet will be attached to the database later on and will not be part of the MasterDataSet. For more information about attaching the information to your database see Part 4, Section 3.

---

### **Technical Assistance**

The WHO Geneva STEPS team will provide technical support during this process if needed. Please contact the team at [steps@who.int](mailto:steps@who.int) with your questions.

---



## Section 3: Data Analysis

### Overview

---

**Introduction** This section covers the tasks that need to be completed to analyse the STEPS survey data. The results of the analysis will be presented in the Fact Sheet and Data Book, which will be used to create the site report.

---

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

- data analyst
- statistical adviser
- STEPS Site Coordinator.

---

**Statistical adviser** If the data analyst is not a survey statistician, it is important that he/she has access to a survey statistician for advice and support. The statistician should be a member of the STEPS Coordinating Committee and have regular contact with the data analyst.

If there is not a statistician available or further assistance is required please contact the WHO Geneva STEPS team at [steps@who.int](mailto:steps@who.int).

---

**Analysis reports** The following reports are the key outputs of the data analysis:

- Data Book
- Fact Sheet
- site report.

---

**Timeframes for analysis** The table below is a guide to when specific parts of the analysis process should begin.

When...	Then...
The data entry templates have been tested.	Begin tailoring the Epi Info code to match your site instrument.
The data is all entered, checked and edited.	Finalize dataset and analyses for the Fact Sheet, main site report, and Data Book.

---

*Continued on next page*

## Overview, Continued

---

### Data analysis software

WHO STEPS recommends using Epi Info for data analysis (version 3.3 or higher), supplemented by Microsoft Access.

Other software packages that are available to the data analysis team may be considered for statistical analyses. However, any alternative packages must be able to handle complex sample designs and will not necessarily be supported by the WHO Geneva STEPS team.

---

### Additional Resources

In addition to the brief introduction to Epi Info presented in Part 3, Section 6 of this manual, there is an Epi Info Training Guide available on the STEPS website and CD that provides more in-depth training.

If you wish to gain a better understanding of the Epi Info analysis programs provided by the STEPS team or you would like to translate these programs into a different statistical package, the Analysis Programs Documentation file is available on the STEPS CD and on the STEPS website here:

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

---

### Technical support

The WHO Geneva STEPS team can provide technical assistance and training for Epi Info to aid the data analyst in the cleaning, weighting, and analysis of the data.

---

### Tasks and timeframes

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

Task Name	Duration	Month 5	Month 6
Prepare and clean the data	7 days		
Calculate response proportions & weight data	7 days		
Create fact sheet and data book	2 weeks		

---

### In this section

This section covers the following topics:

Topic	See Page
Data Analysis Process	4-3-3
Preparing the Survey Data	4-3-4
Cleaning the Data	4-3-7
Calculating Response Proportions	4-3-12
Weighting the Data	4-3-13
Finalizing the Dataset	4-3-19

---

## Data Analysis Process

---

### Introduction

The data analysis process ranges from creating the database to producing the final results for the site report.

Data analysis should be conducted in a very standardized and methodical way, using the guidelines suggested by the STEPS team. Standardizing certain aspects of the data analysis will allow trend analysis in the future between STEPS surveys and also allow comparisons between STEPS sites.

---

### Process

The table below shows each of the stages in the data analysis process.

Stage	Description
1	Preparing the STEPS data for Epi Info
2	Cleaning the data
3	Weighting the data
4	Producing the Fact Sheet and Data Book

**Note:** Stages 1-3 are described in this section, stage 4 is described in the following section of the manual.

---

## Preparing the Survey Data

---

### Introduction

Once data entry is complete, the data needs to be prepared for cleaning and analysis in Epi Info and the interview tracking information needs to be attached to the survey data. Note that much of this topic is a review of material covered in Part 2, Section 5 of this manual.

---

### Instructions for importing data from EpiData

Upon completion of the data entry process, the data entry supervisor will produce a single database file (.dbf) containing all the entered data from the instruments. This .dbf file must be transformed into a data table called MasterDataSet in an Access file called STEPS.mdb in order for the provided analysis programs to work properly. The table below describes how to transform your data from the .dbf file to the Access database file.

Step	Action
1	Rename the .dbf file containing your survey data to "STEPS.dbf".
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\Epi Info folder.

---

### Epi Info 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 following table describes how to download the Epi Info analysis programs from the internet and attach them to your STEPS.mdb file containing your survey data.

---

*Continued on next page*

## Preparing the Survey Data, Continued

### Epi Info analysis programs (cont.)

Step	Action
1	Connect to the internet and go to: <a href="http://www.who.int/chp/steps/resources/database/en/index.html">http://www.who.int/chp/steps/resources/database/en/index.html</a>
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.
5	Open the Epi_Info_Analysis_Programs.mdb file and right click on the table "Programs".
6	Select "Export" and in the "Select Table to Export to" window select STEPS.mdb (this should be located in your C:\STEPS\Epi Info folder).
7	Select the "Definition and Data" option in the dialogue window and click "OK". If you are replacing an older version of the programs, a dialogue window will appear asking if you want to replace the old "Programs" table. Click "Yes".

### Import interview tracking form

If interview tracking information has been collected and entered into the interview\_tracking\_form.xls Excel file, import the information contained in this file into your STEPS.mdb file by following the steps below.

Step	Action
1	Confirm that the interview_tracking_form.xls file is located in your C:\STEPS\data folder and that it is complete and ready to be imported into Epi Info (see Part 2, Section 5 for more detailed instructions).
2	Open Epi Info Analysis and click on Analysis Commands>User-Defined Commands>Run Saved Program in the Analysis window.
3	Click the grey box to the right of the Filename field and find and select your STEPS.mdb file.
4	Select "ImportInterviewTracking" from the drop down menu and click "OK".

**Note:** This process uses one of the generic programs written for Epi Info and therefore must be done after the programs have been attached to your dataset.

*Continued on next page*

# Preparing the Survey Data, Continued

---

**Create backup of database**

It is important to create a backup of your database. During the analysis process you will be writing and saving different tables within your database. If something happens to your working copy of the database you will need a backup copy. Follow the steps below to create a backup of your database.

<b>Step</b>	<b>Action</b>
1	Open STEPS.mdb.
2	From the File menu click on "Back up Database".
3	Select a location on your machine to back up the database.
4	Click "Save".

---

## Cleaning the Data

### Introduction

The dataset needs to be cleaned prior to data analysis. This includes:

- checking ranges and combinations of variables
- detecting and handling missing data
- detecting and handling outliers.

### Checking Age and Sex Variables

The variables Age and Sex should be checked first, prior to checking the data in any other variable. Age and Sex are needed in order to analyse the survey data by age-sex groups and can also be useful in cleaning the remaining variables.

There are a set of programs available in the provided Epi Info analysis programs that help to automate the process of checking the Age and Sex variables. These are:

- AgeRange2564 (or AgeRange1564)
- Rerun\_AgeRange2564 (or Rerun\_AgeRange1564)
- MissingAgeSex

The table below explains how to use these programs to check your age and sex variables.

Step	Action								
1	Open Epi Info Analysis and click on Analysis Commands>User-Defined Commands>Run Saved Program in the Analysis window.								
2	Click the grey box to the right of the Filename field and find and select your STEPS.mdb file.								
3	Select "AgeRange2564" from the drop-down list in the Program field and click "OK".								
4	<table border="1"><thead><tr><th>If the program result is...</th><th>Then...</th></tr></thead><tbody><tr><td>There are no records missing age or sex</td><td>Run MissingAgeSex</td></tr><tr><td>There are records missing age or sex that cannot be resolved</td><td>Run MissingAgeSex</td></tr><tr><td>There are records missing age or sex that can be resolved</td><td><ul style="list-style-type: none"><li>• Resolve records</li><li>• Run Rerun_AgeRange2564</li><li>• Run MissingAgeSex</li></ul></td></tr></tbody></table>	If the program result is...	Then...	There are no records missing age or sex	Run MissingAgeSex	There are records missing age or sex that cannot be resolved	Run MissingAgeSex	There are records missing age or sex that can be resolved	<ul style="list-style-type: none"><li>• Resolve records</li><li>• Run Rerun_AgeRange2564</li><li>• Run MissingAgeSex</li></ul>
If the program result is...	Then...								
There are no records missing age or sex	Run MissingAgeSex								
There are records missing age or sex that cannot be resolved	Run MissingAgeSex								
There are records missing age or sex that can be resolved	<ul style="list-style-type: none"><li>• Resolve records</li><li>• Run Rerun_AgeRange2564</li><li>• Run MissingAgeSex</li></ul>								

*Continued on next page*

## Cleaning the Data, Continued

### Checking Age and Sex Variables (cont.)

Step	Action
5	<p>Upon completion of this process, your STEPS dataset will contain the following new variables:</p> <ul style="list-style-type: none"> <li>• AgeRange, containing text values "25-34", "35-44", etc.</li> <li>• Age, containing numeric age values</li> <li>• Sex, containing text values "Men" and "Women"</li> <li>• Valid, containing numeric values 1 (valid) and 2 (not valid)</li> </ul> <p>In order to have a value of 1 (valid) for the variable Valid, a record must have a valid Age and Sex value.</p>

**Note:** If the overall age range for your survey is neither 15-64 nor 25-64, or you have not planned to report estimates for 10-year age groups, contact the STEPS team for help in modifying the AgeRange2564 and Rerun\_AgeSex2564 programs.

### Checking variables needed for weighting

If data will be weighted for probability of selection, variables indicating the location of each record (e.g. Cluster Number) must be checked for missing or outlying values. As the name and value of these variables vary from country to country, no automated program exists for this cleaning.

The table below is provided as a general guideline to check the location variables in your dataset, please contact the STEPS team for more specific help with checking your location variables.

Step	Action
1	Consult with the data entry supervisor to identify all variables in the dataset containing location information and the accepted range of values for each.
2	For each variable, confirm that all records contain data within the accepted range of values. (In Epi Info, the LIST and FREQ (frequency) commands may be useful. See Part 3, Section 6 for help with Epi Info commands.)
3	Confirm that location information is consistent within each record (In Epi Info, the SELECT and FREQ (frequency) commands may be useful.)
4	If missing or inconsistent data is encountered while checking the location variables, contact the data entry supervisor to discuss the possibility of correcting the errors. If the data cannot be corrected, the record will not be able to be included in the weighted analyses.

*Continued on next page*

## Cleaning the Data, Continued

---

### Automated cleaning

There is some basic cleaning code embedded within most of the provided Epi Info analysis programs which will clean the data for:

- basic outliers;
- completeness;
- consistency (e.g. if a participant said No to currently smoking and then Yes to smoking daily).

For each program, records with outlying data or incomplete or inconsistent responses (where more than one question is needed for analysis) are temporarily removed during the analysis.

Even though this basic cleaning is incorporated into most of the provided programs, it is still highly recommended that each variable and set of variables be checked for outliers, completeness, and consistency prior to any analyses.

---

### Missing data

While most problems related to missing data should be handled during the data entry process (see Part 4, Section 2), it is still very likely that the final dataset will contain records with missing data. Thus, the data analyst will need to explore the missing data in greater depth so that the data analysis is properly completed with attention to this missing data.

---

### Guidelines for handling missing data

In general, how missing data is handled depends upon the importance of the variable and how much data is missing. Use the table below when considering how to handle missing data in the following situations.

If...	Then...
A record is missing data for age, sex, or any location variable (if data are weighted).	Review the completed instrument and discuss with the data entry supervisor to try to recover the missing data. If the data cannot be recovered, the record should be dropped and counted as a non-responder for weighting purposes.
A record is missing data for a variable other than age, sex or location.	Exclude the record from all analyses relating to this variable (the automated Epi Info programs will do this automatically). For analyses in which records are excluded, include a footnote stating the number of records omitted due to missing data.

---

*Continued on next page*

## Cleaning the Data, Continued

---

**Imputation** An alternative method of handling missing data, that ‘creates’ data where none exists, is called imputation. Imputation should **not** be done for STEPS.

---

**Outliers** An outlier is a value of a variable that appears to deviate significantly from the observed values in other participants. It may be correct, and the person truly has an unusual value, or it may be incorrectly recorded or entered. In any case, it is good practice to investigate the outliers before analysis in order to avoid having those extreme values unduly influencing the results being reported.

---

**Identifying outliers** The data entry templates provided for EpiData do not allow data to be entered outside of an expected range of values for nearly all questions on the STEPS Instrument. Therefore using the provided templates should minimize the presence of outliers in the dataset.

Most Epi Info analysis programs also ensure that the values for a given variable of interest fall within an expected range. The value ranges used in the Epi Info programs match those used in the EpiData templates. These values are listed in the table below.

Variable Description	Standard Variable Code	Accepted Values
Age, in years	C3	Age range of survey
Years of education	C4	0-30
Number of people $\geq$ 18 yrs. in household	C9	1-30
Age started/stopped smoking daily	T3; T7	10-74
Time since starting/stopping smoking daily	T4a-c; T8a-c	1-64 years 1-12 months 1-30 days
Number of tobacco products smoked/used each day	T5a-e; T11a-e	0-50
Number of occasions alcohol consumed	A4	1-50
Number of drinks consumed on given occasion	A5; A6	1-50
Number of occasions alcohol consumed in large quantities	A7	0-50
Number of drinks consumed per day	A9a-g	0-50
Number of servings of fruit or vegetables consumed on a given day	D2; D4	1-20

*Continued on next page*

## Cleaning the Data, Continued

### Identifying outliers (cont.)

Variable Description	Standard Variable Code	Accepted Values
Number of meals eaten outside the home	D6	0-30
Amount of physical activity per day	P3a-P3b; P6a-P6b; P9a-P9b; P12a-P12b; P15a-P15b	00:10-16:00
Amount of sedentary activity per day	P16a-P16b	00:00-24:00
Height	M3	100-270 cm
Weight	M4	20-350 kg
Waist circumference	M7	30-200 cm
Systolic blood pressure	M11a; M12a; M13a	40-300 mmHg
Diastolic blood pressure	M11b; M12b; M13b	30-200 mmHg
Hip circumference	M15	45-300 cm
Fasting blood glucose	B5	1-35 mmol/l 18.0-630.0 mg/dl
Total cholesterol	B8	1.75-20.00 mmol/l 67.0-773.0 mg/dl
Fasting triglycerides	B10	0.25-50.00 mmol/l 22.0-4428.0 mg/dl
HDL cholesterol	B11	0.1-5.0 mmol/l 3.8-190.0 mg/dl

If the EpiData templates and/or Epi Info analysis programs are not used, each variable should be checked for outliers before any analysis is completed. The table below provides some suggestions on how to check for outliers in Epi Info. More information about these commands are provided in Part 3, Section 6 of this manual and in the STEPS Epi Info Training Guide available on the STEPS website.

Command Name	Use
FREQUENCY	To list the frequency of all values for a categorical or continuous variable.
MEANS	To list the frequency of values, the mean value, the standard deviation, and the quartile values of a continuous variable.
GRAPH	To obtain a visual representation (e.g. histogram or scatter plot) of all values for a categorical or continuous variable.

# Calculating Response Proportions

---

## Introduction

Response proportions (often known as response rates) indicate the level of participation in your STEPS survey. They are an important indicator of the quality of your data.

---

## Interview Tracking Form

Response proportions are calculated from the information entered in the Interview Tracking Form. This form should already have been:

- entered into the interview\_tracking\_form.xls file by the data entry team (see Part 4, Section 2);
- imported into Epi Info (see page 4-3-5).

Since response rates for Step 3 typically differ from those of Step 1 and Step 2, the generic programs compute the response rate for Step 3 separately (if applicable for your survey). If the response rate for your survey differs markedly between Step 1 and Step 2, contact the STEPS team for assistance in computing separate response rates for these Steps.

---

## Calculating response proportions in Epi Info

Follow the steps below to run the Epi Info Programs necessary to calculate the response proportions for each Step.

Step	Action
1	Open Epi Info Analysis and click on Analysis Commands>User-Defined Commands>Run Saved Program in the Analysis window.
2	Click the grey box to the right of the Filename field and find and select your STEPS.mdb file.
3	Select "ResponseOverall" from the drop down menu and click "OK".
4	Repeat the above steps to run the program "ResponseStep3" if appropriate.

---

## Weighting the Data

---

**Introduction** The data from your STEPS survey only represents the participants sampled. If you want your data to be representative of the target population then you will need to apply weights to your data.

---

**What is a weight** A weight is a value given to a data record to adjust the importance given to it in analysis. It may be thought of as the number of persons in the population that are represented by each individual in the sampled unit. Weights are calculated to adjust for the following aspects of a survey:

- probability of selection (sample weight);
  - non-response (non-response weight);
  - differences between the sample population and target population (population weight).
- 

**Types of weights** The table below lists the 3 different types of weights and where the information for these weights comes from.

Type of Weight	Used to...	Required when ...	Information available in...
Sample weight	adjust for the probability of selection of each participant.	only a sample of the population is selected for the survey.	STEPSsampling.xls and interview_tracking_form.xls
Non-response weight	adjust for differential response proportions.	there is a low response rate for your survey overall or within a specific age-sex group.	interview_tracking_form.xls
Population weight	adjust for differences in the sample vs. target population in age-sex composition.	the sample vs. target population differ markedly in age-sex composition.	STEPSsampling.xls

---

**Overall weight** All the weights described in the table above are calculated for each record and multiplied together. Using the abbreviations of w1 for sample weight, w2 for non-response weight, and w3 for population weight, the overall weight would be calculated as follows:

$$\text{Overall weight} = w1 * w2 * w3$$

---

*Continued on next page*

## Weighting the Data, Continued

**Overall weight (cont.)** If differences exist in the response rates for each Step or if a subsample is used for Step 3, then a separate overall weight will need to be calculated for each Step of the survey. If, for example, only a subsample of the overall sample participated in Step 3, then the overall weights would be calculated as follows:

Overall weight for Step 1 & 2 =  $w_{1Step12} * w_{2Step12} * w_{3Step12}$

Overall weight for Step 3 =  $w_{1Step3} * w_{2Step3} * w_{3Step3}$

**Stratum and PSU** If your sample design was anything other than a simple random sample, you will need to create variables that contain information about your sample design. These variables are conventionally named Stratum and PSU and their values depend on the sample design of your survey. These variables should be calculated and attached to your dataset **before** calculating and attaching the weights.

The table below explains what the values of Stratum and PSU should be for 3 common sample designs (for more information about sample designs, see Part 2, Section 2), if you do not see your sample design listed here or would like assistance in determining the values of your Stratum and PSU variables, contact the STEPS team.

Sample Design	Stratum Values	PSU Values
Stratified multi-stage cluster sampling	1 unique value for each stratum	1 unique value for each first-stage sampling unit within each stratum
Multi-stage sampling without stratification*	1 unique value for each first-stage sampling unit	1 unique value for each second-stage sampling unit within each first-stage sampling unit
Two-stage sampling without stratification.	Stratum not needed	1 unique value for each first-stage sampling unit

**\*Note:** Refer to the instructions in the "Info for Weighting" worksheet in STEPSsampling.xls for more detailed directions on assigning Stratum and PSU values for this type of sample design.

**Sample weight** The sample weight is the most difficult weight to calculate due to the amount of information needed. While there are some tools available to help with the calculation of these weights, it is not possible to automate the process entirely due to differences in sample design between STEPS surveys.

*Continued on next page*

## Weighting the Data, Continued

---

### Sample weight (cont.)

If you used the STEPSsampling.xls file to draw your sample, you can use it to partially calculate the sample weights for your dataset. The worksheet "Info for Weighting" within the STEPSsampling.xls file contains directions for calculating the probability of selection up to the household or individual level (if individuals were selected directly).

If individuals were selected using the Kish Method (recommended) or other within-household sampling method, you will need to also calculate the probability of selection for each participant within the household. This information can be found in the completed interview\_tracking\_form.xls file, which contains the participant ID along with the number of eligible individuals in each participant's household.

To calculate the final sample weight for each participant, follow the steps below:

Step	Action
1	Calculate the probability of selection at all stages of sampling before the household level.
2	Multiply together the probabilities calculated in step 1 above. This product should be the same for participants selected from the same sampling units (e.g. districts, villages).
3	Calculate the within-household probability of selection for each participant and multiply this by the product obtained in step 2 above.
4	Take the inverse of the product obtained in step 3 above to obtain the final sample weight.

### Non-response weight

For STEPS, it is recommended to examine the non-response rate for each age-sex group. If one or more age-sex groups have a low response rate, a non-response weight should be calculated for each age-sex group and applied accordingly to all records in the dataset.

The non-response weight is simply the inverse of the response rate for each age-sex group. It can be calculated automatically in Epi Info if the interview\_tracking\_form.xls file has been completed. If you have not used the interview\_tracking\_form.xls file, contact the STEPS team for assistance in calculating and attaching the non-response weights for your data.

---

*Continued on next page*

## Weighting the Data, Continued

### Non-response weight (cont.)

To use the provided Epi Info program to automatically calculate the non-response weight and attach the weight to the dataset, follow the steps below.

Step	Action
1	Ensure that the interview_tracking_form.xls file has been attached to your STEPS database (see 4-3-5 for instructions).
2	Open Epi Info Analysis and click on Analysis Commands>User-Defined Commands>Run Saved Program in the Analysis window.
3	Click the grey box to the right of the Filename field and find and select your STEPS.mdb file.
4	Select "NonresponseWeight" from the drop down menu and click "OK". The program will generate the variables W2s1, W2s2, and W2s3 which contain the non-response weight for Step 1, 2, and 3, respectively.

**Note:** The non-response weights calculated by the Epi Info program do not take into consideration records that may be excluded due to missing location variables or due to non-fasting for Step 3. Therefore, these weights may need adjustment to take into consideration any records excluded for these reasons.

### Population weight

Prior to calculating the population weight, you must first check to see if there are marked differences between the age-sex structure of your sample and the age-sex structure of your target population. It is possible, for example, that your sample has a much smaller proportion of men in the 55-64 year age group than your target population does. The population weight will help to correct for such differences.

To compute the proportion of your sample and target population comprised of each age-sex group of interest, follow the steps below.

Step	Action
1	Calculate the total number of individuals in your sample.
2	Calculate the total number of individuals in each 10-year age-sex group in your sample.
3	Divide the total number of individuals in each age-sex group by the total number of individuals in your sample to obtain the proportion of your sample comprised of each age-sex group.
4	Repeat steps 1-3 above for the target population to obtain the proportion of your target population comprised of each 10-year age-sex group.
5	For each 10-year age-sex group, compare the proportion in the sample to that in the target population. If there are marked differences in any age-sex group, population weights should be calculated for your dataset.

*Continued on next page*

## Weighting the Data, Continued

---

### Population weight (cont.)

If the above calculations reveal that population weights are needed for your dataset, the weights can be calculated for each 10-year age-sex group by dividing its proportion in the total population by its proportion in the target population. There is an Epi Info program that will do this calculation for you automatically and attach the population weight to your dataset.

Follow the steps below to calculate population weights and attach them to your dataset.

Step	Action
1	Complete the "Population Info" spreadsheet in the STEPSsampling.xls file. Be sure to keep a record of the data source for your target population (e.g. 2005 Census) as this information is needed for the site report.
2	Click on the "Format for Epi Info" button on this spreadsheet to generate the "PopulationEst" spreadsheet that will be read by Epi Info.
3	Close your STEPSsampling.xls file and ensure that it is located in your C:\STEPS\data folder.
4	Open Epi Info Analysis and click on Analysis Commands>User-Defined Commands>Run Saved Program in the Analysis window.
5	Click the grey box to the right of the Filename field and find and select your STEPS.mdb file.
6	Select "PopulationWeight" from the drop down menu and click "OK". The program will generate the population weight variable W3.

**Note:** The EpiInfo program calculates the population weight using all valid records in the dataset. If a only a subset of your sample participated in Step 3, then a separate Step 3 population weight should be calculated.

---

### Overall weights

Once the sample, non-response, and population weights have been calculated and attached to your dataset (as needed), you will need to multiply these together to arrive at the overall weight for each Step of your survey. The Epi Info weighted analysis programs expect the overall weight for each Step to be named:

- WStep1
- WStep2
- WStep3

---

*Continued on next page*

## Weighting the Data, Continued

---

### Calculating overall weights (cont.)

Follow the steps below to calculate the overall weights in Epi Info once you have attached the sample, non-response, and population weights to your dataset. It may help to refer to Part 3, Section 6 for more information about the Epi Info commands used.

Step	Action
1	Open Epi Info Analysis and click on Analysis Commands>Data>Read (Import) in the Analysis window.
2	Find and select your MasterDataSet data table in your STESP.mdb file.
3	Use the DEFINE command to create the variable "WStep1".
4	Use the ASSIGN command to set "WStep1" equal to the product of the sample, non-response, and population weights (as applicable) calculated for Step 1.
5	Repeat steps 3-4 to define and assign variables "WStep2" and "WStep3" as needed.
6	Save your dataset by going to Analysis Commands>Data>Write (Export).

---

## Finalizing the Dataset

---

### Introduction

Before proceeding with any analyses, consult the table below to ensure that your dataset has been properly prepared for unweighted or weighted analysis.

Step	Action
1	Ensure your database is named STEPS.mdb and is located in your C:\STEPS\EpiInfo folder.
2	Ensure that your dataset within your STEPS.mdb file is called "MasterDataSet".
3	Ensure that the variables Age, Sex, AgeRange, and Valid have been created using the AgeRange2564 (or AgeRange1564) and MissingAgeSex programs.
4	Ensure any outliers and missing data cannot be corrected or recovered from the original instruments.
5	If weighted analyses are to be done, ensure the variables Stratum, PSU, WStep1, WStep2, and WStep3 (as applicable) have been created in your dataset.

---

### Next step

Once all you have confirmed that all of the above steps have been completed, you may proceed to create the Fact Sheet and Data Book. Continue to the next section, Part 4, Section 4, for more information about this process.

---



## Section 4: Reporting and Disseminating Results

### Overview

---

<b>Introduction</b>	This section covers the tasks that are needed to prepare reports and disseminate the results of your STEPS survey.
<b>Requirement</b>	<p>The reports need to be produced in a timely manner after the completion of your survey. The results should be presented in a clear, concise and usable way to help:</p> <ul style="list-style-type: none"><li>• raise awareness about preventing chronic disease and their risk factors</li><li>• guide public health policy and interventions to address chronic diseases</li><li>• assist and inform future health research.</li></ul>
<b>Intended audience</b>	<p>This section is primarily designed to be used by those fulfilling the following roles:</p> <ul style="list-style-type: none"><li>• STEPS Site Coordinator</li><li>• data analyst</li><li>• STEPS Coordinating Committee.</li></ul>
<b>Useful resources</b>	<p>Some sections of the manual that may be useful in compiling and disseminating the results include:</p> <ul style="list-style-type: none"><li>• Part 1, Section 1 : "Introduction";</li><li>• Part 2, Section 2 : "Preparing the Sample";</li><li>• Part 6, Section 3A-E: "Report Templates" (includes Fact Sheet, Data Book, Site Report Template);</li><li>• Part 7, Section 1 : "Glossary of Terms Used in STEPS".</li></ul>
<b>PowerPoint Presentation</b>	<p>There is a useful PowerPoint presentation available on the STEPS CD and website (<a href="http://www.who.int/chp/steps">www.who.int/chp/steps</a>) that provides information on interpreting data and disseminating your results. It is advised that you look at this presentation prior to writing your site report. The PowerPoint presentation includes information on:</p> <ul style="list-style-type: none"><li>• summarizing data in a meaningful way;</li><li>• making good graphics (including tables, bar graphs, line graphs and pie charts);</li><li>• interpretation of results;</li><li>• confidence intervals and standard error;</li><li>• using confidence intervals to test for subgroup differences.</li></ul>

---

*Continued on next page*

## Overview, Continued

### The reports

The main reporting documents that need to be produced are listed in the table below. More information for each reporting document is also provided in this section further below.

<b>Report</b>	<b>Brief description</b>	<b>Audience</b>
Fact Sheet	Short summary of key results	<ul style="list-style-type: none"><li>• Stakeholders</li><li>• Media</li><li>• STEPS teams</li></ul>
Data Book	Detailed results presented in generic tables	<ul style="list-style-type: none"><li>• STEPS teams</li></ul>
Site report	Main comprehensive report	<ul style="list-style-type: none"><li>• Stakeholders</li><li>• Media</li><li>• Wider community</li><li>• STEPS teams</li></ul>
Progress report (optional)	Report on surveillance progress	<ul style="list-style-type: none"><li>• Site Coordinator</li><li>• Coordinating Committee</li><li>• WHO Geneva STEPS team</li><li>• WHO Regional Office</li></ul>

### Reporting process

The table below shows each of the key stages in the reporting process once data have been entered, checked and analysed.

<b>Stage</b>	<b>Description</b>
1	Preparing and distributing the Fact Sheet to cover the essential results.
2	Preparing the Data Book.
3	Extracting specific tables from the Data Book that are suitably weighted and needed for the main site report.
4	Drafting the main site report, section by section, based on the content guidelines (see Part 6, Section 3E) and Data Book.
5	Circulating drafts of the site report to members of the coordinating committee, WHO and other interested parties for comment, discussion and review.
6	Reviewing and finalizing the site report in light of comments and discussions.
7	Preparing circulation lists, preparing press releases and promotion fliers to announce results of the STEPS Survey.
8	Presenting results, through slide presentations and meetings with organizations and groups that have an interest and impact on population health including relevant government departments, sponsors, tertiary institutions and health conferences in order to widen awareness of the STEPS findings.

*Continued on next page*

## Overview, Continued

---

**In this section** This section covers the following topics.

<b>Topic</b>	<b>See Page</b>
Summarizing Data	4-4-4
Displaying Results	4-4-7
Interpretation of Results	4-4-8
Preparing and Distributing the Fact Sheet	4-4-9
Preparing and Distributing the Data Book	4-4-12
Preparing and Distributing the Site Report	4-4-14
Progress Report	4-4-15

---

## Summarizing Data

---

**Introduction** STEPS data on chronic disease risk factors that have been collected from individuals need to be summarized in a meaningful way in order to give relevant information on levels of risk factors in a population.

Summary statistics that are used for summarizing STEPS data include:

- mean
- median
- prevalence.

When using the STEPS EpiInfo Programs, your output tables will display these summary statistics. The three summary statistics are described in more detail below.

---

**Mean** The mean is a measure of central tendency and is computed by adding all the individual values in the group and dividing by the number of values in the group. It gives information on a population's average of a specified variable, such as waist circumference or blood sugar level (96).

---

**Median** The median is another measure of central tendency that is often used for non-normally distributed variables. It is the simplest division of a set of sorted measurements into two halves - the lower and the upper half (96). The median is often reported along with the 25<sup>th</sup> and 75<sup>th</sup> percentiles, which are the values that separate the lowest 25% and highest 75% of values, respectively, in the set of measurements (96).

---

**Prevalence** Prevalence is defined as the number of persons with a disease or an attribute in a given population at a designated time, e.g. % daily smoker in a country in 2008 (96).

---

**Standard cut-offs for prevalence** In order to determine the prevalence of those persons in a specified population that are at risk to develop a chronic disease, cut-off points have been set up to distinguish between "at risk" and "not at risk". STEPS uses those cut-offs that are evidence-based, widely used and therefore recommended by the World Health Organization. See below the standard cut-offs STEPS uses for the eight main risk factors.

---

**Cut-offs for tobacco use** The two main tobacco indicators that are associated with an increased risk of developing chronic diseases are

- current use of tobacco
  - daily use of tobacco (97,98).
- 

*Continued on next page*

## Summarizing Data, Continued

---

### Cut-offs for alcohol consumption

Five main indicators should be reported for alcohol consumption, including

- lifetime abstainer;
  - last year abstainer;
  - hazardous drinking, defined as consuming 40-59.9g of pure alcohol on average per day for men, and 20-39.9g for women;
  - harmful drinking, defined as consuming  $\geq 60$ g of pure alcohol on average per day for men, and  $\geq 40$ g for women;
  - binge drinking, defined as drinking  $\geq 5$  drinks in a row for men, and  $\geq 4$  drinks in a row for women (18, 99, 100).
- 

### Cut-offs for fruit and vegetable intake

Eating less than five servings of fruit and/or vegetables per day is considered being a low fruit and vegetable intake and increases the risk to develop chronic diseases (46).

---

### Cut-offs for physical activity

A person not meeting any of the following criteria is considered being physically inactive and therefore at risk of chronic disease:

- 3 or more days of vigorous-intensity activity of at least 20 minutes per day; OR
  - 5 or more days of moderate-intensity activity or walking of at least 30 minutes per day; OR
  - 5 or more days of any combination of walking, moderate- or vigorous-intensity activities achieving a minimum of at least 600 MET-minutes per week (101-103).
- 

### Cut-offs for obesity and overweight

The body mass index (BMI) is a statistical measure of the weight of a person scaled according to height. It is defined as the individual's body weight divided by the square of the height ( $\text{kg}/\text{m}^2$ ).

A BMI  $\geq 25$  indicates that a person is overweight, while a BMI  $\geq 30$  indicates that the person is obese (57, 104-107).

---

### Cut-offs for raised blood pressure

Blood pressure is commonly measured in mmHg. A person is considered being mildly hypertensive if the systolic value (SBP)  $\geq 140$  mmHg and/or the diastolic value (DBP)  $\geq 90$  mmHg.

Moderate hypertension has been defined as SBP  $\geq 160$  mmHg and/or DBP  $\geq 100$  mmHg (108).

---

*Continued on next page*

## Summarizing Data, Continued

---

### Cut-offs for raised blood glucose

Blood glucose is commonly measured after fasting. It can be measured either by determining the glucose value in the venous blood, or in the capillary blood. The corresponding cut-offs are slightly different. Raised blood glucose has been defined as

- plasma venous value  $\geq 7$  mmol/l (equivalent to  $\geq 126$  mg/dl)
  - capillary whole blood value  $\geq 6.1$  mmol/l (equivalent to  $\geq 110$  mg/dl) (109).
- 

### Cut-offs for abnormal blood lipids

The two main lipids in the blood are cholesterol and triglycerides and therefore, different indicators should be reported for blood lipids:

- total blood cholesterol  $\geq 190$  mg/dl (equivalent to 5.0 mmol/L);
  - total blood cholesterol  $\geq 240$  mg/dl (equivalent to 6.2 mmol/L);
  - HDL cholesterol  $< 50$  mg/dl for men and  $< 40$  mg/dl for women (equivalent to 1.3 mmol/L for men and 1.0 mmol/L for women);
  - triglycerides  $\geq 180$  mg/dl (equivalent to 2.0 mmol/L);
  - triglycerides  $\geq 150$  mg/dl (equivalent to 1.7 mmol/L) (110-113).
- 

### Estimates and true values

Prevalence, mean and median are estimated values, as they usually derive from a sample, and not from the target population (for more on sampling, see Part 2, Section 2). In order to give information on how uncertain estimated values are, confidence intervals are computed around the estimate.

---

### Standard error and Confidence Interval (CI)

A standard error is the standard deviation of an estimate, e.g. a mean. It can be used to calculate confidence intervals.

A confidence interval is a computed interval with a given probability, e.g. 95%, that the true value of a variable such as a mean or a prevalence is contained within the interval (96).

---

## Displaying Results

---

### Introduction

Results can be displayed by either describing summary statistics such as prevalence, mean or median in words, or by using tables and graphs. Sometimes, visual methods can make the point much stronger than simply describing the data. However, they need to be prepared with sufficient care. See the guidelines below for making good tables and graphs.

---

### Guidelines for making good tables and graphs

The general guidelines below may help when preparing tables and graphs.

- Each table or graph should contain enough information so that it can be interpreted without reference to the text.
  - Titles of tables and graphs should specifically describe the numbers included.
  - Decide on the point you wish to present, then choose the appropriate method.
  - Specify the units being used clearly.
- 

### Guidelines for making good tables

See guidelines below for making good tables.

- Tables should always include, for all age groups, the number of respondents / the denominator for which the table is made.
- The total age group should be highlighted.
- Confidence intervals, if available, should be included in tables.
- Vertical lines in tables should be avoided.

See an example of a prevalence table below.

Example table 1. Percentage of daily smoker in country x, both sexes, 2006.

Age group	n	% daily smoker	95% CI
25-34	1000	20.0	18.0-22.0
35-44	1500	19.5	18.5-20.5
45-54	1500	23.6	22.6-24.6
55-64	1000	15.4	13.4-17.4
<b>25-64</b>	<b>5000</b>	<b>19.7</b>	<b>19.2-20.2</b>

---

### Guidelines for making good graphs

See guidelines below for making good graphs.

- Mention the number of respondents for which the graph is made.
  - Be sparing and consistent with use of colors and fonts.
  - Emphasize one idea at a time in a graph.
  - Examples of graphs include pie charts, bar charts or line graphs.
-

## Interpretation of Results

---

### **Introduction**

In order to deliver a meaningful message, results need to be interpreted carefully. Influence factors such as sample size, response percentages, season of data collection or potential biases need to be thought through and taken into account when interpreting results. Below are a few bullet points that help when interpreting results.

---

### **Representativeness of results**

Results should only be applied to the surveyed target population, and not be generalized to a broader population.

---

### **Uncertainty of results**

Confidence intervals help to determine the uncertainty of the estimates. The shorter the interval the better (*114*).

---

### **Influence on results**

Think through carefully what could have influenced the results when interpreting them. Potential influence factors include:

- sample sizes (Are they high enough to have produced robust results for all subgroups?);
  - response rates (Are they high or low? Are they the same for all subgroups, or have some subgroups lower response rates than others? If so, why?);
  - social pressure (May people have answered in a specific way to certain questions because of social desirability?);
  - survey methodology (Could flaws/problems in survey methods have influenced results, e.g. problems in reaching working population during data collection?);
  - participant comprehension (Are there specific questions in the questionnaire that seemed not to be understood by respondents?);
  - season of data collection.
- 

### **Results in a context**

When interpreting results, it is useful to put results in a context. As an example, you may want to find out about the amount of cigarettes being sold when looking into results for prevalence of cigarette smokers in a country.

---

## Preparing and Distributing the Fact Sheet

---

<b>Introduction</b>	The Fact Sheet is a short summary of the key results of the STEPS chronic disease risk factor survey.
<b>Purpose</b>	The purpose of the Fact Sheet is to provide interested parties with the key findings of the survey and to highlight the issues that the main report will cover in more depth.
<b>Intended audience</b>	<p>It is recommended that you distribute the Fact Sheet widely. Consider sending copies to:</p> <ul style="list-style-type: none"><li>• relevant government bodies and sponsoring organizations;</li><li>• agencies and organizations who are likely to use the information to promote healthy lifestyles or determine health policies;</li><li>• public, governmental and institutional (university) libraries;</li><li>• press and other media (newspapers, radio and television);</li><li>• websites dealing with chronic diseases and related health issues;</li><li>• WHO Regional Office, STEPS focal point and the WHO Geneva STEPS team.</li></ul>
<b>Standardized format and results</b>	<p>A STEPS Fact Sheet Template is available in Part 6, Section 3C. Use this template to present the summarized results in a standardized format.</p> <p>If you have added optional questions that are not represented in this template, only include these in the Fact Sheet if they are particularly significant.</p>
<b>Fact Sheet layout</b>	<p>The STEPS Fact Sheet contains a short paragraph that briefly describes when and where the STEPS survey has been carried out, the scope of the survey, as well as age groups covered, overall sample size and response rates, and a short description of the sampling method.</p> <p>The Fact Sheet table contains blocks of results on tobacco use, alcohol consumption, fruit and vegetable consumption, physical activity, physical and biochemical measurements, and a summary of combined risk factors. Lines in this table include brief descriptions of the most important indicators of each risk factor, followed by results columns for both sexes, males and females.</p> <p>The last lines on the Fact Sheet display contact details of the respective STEPS country focal point.</p>

---

*Continued on next page*

## Preparing and Distributing the Fact Sheet, Continued

---

### Fact Sheet Analysis Guide

The Fact Sheet Analysis Guide has been developed to help preparing the Fact Sheet. It looks similar to the STEPS Fact Sheet, but instead of the results columns, it contains one column that displays the standard question code of the questions required to calculate results, and one column that includes the names of the Epi Info Programs that need to be run to produce the results.

---

### Process

The table below lists the processes required to prepare your Fact Sheet and to generate the results for the Fact Sheet after all cleaning of the data has been completed. Generic code for Epi Info has been written to generate all the indicators presented on the Fact Sheet.

Stage	Description
1	Insert site name and year of STEPS survey in the heading of the Fact Sheet
2	Insert the following in the text above the table on the Fact Sheet: <ul style="list-style-type: none"><li>• site name</li><li>• months and year of STEPS survey</li><li>• sampling design</li><li>• sample size</li><li>• response rate</li><li>• year of planned repeat survey.</li></ul> In addition, all necessary adjustments need to be made to the text, such as to indicate whether or not Step 3 has been done.
3	Insert contact details of your STEPS country focal point below the table.
4	Determine which indicators on the Fact Sheet can be computed for your site using a copy of your site instrument and the column "Questions required to calculate result" on the Fact Sheet Analysis Guide.
5	Using the Fact Sheet Analysis Guide, circle the names of the programs associated with the indicators that can be computed.
6	Remove those rows from the Fact Sheet for which results can not be produced because the required questions have not been asked.
7	Run all Epi Info programs circled on the Fact Sheet Analysis Guide (see Part 3, Section 6 for help with running programs in Epi Info).

---

*Continued on next page*

## Preparing and Distributing the Fact Sheet, Continued

### Process (cont.)

Stage	Description
8	<p data-bbox="555 387 1410 600">Copy the corresponding results (including confidence intervals) for both sexes, males and females for all indicators that can be computed from the output of the Epi Info programs into the Fact Sheet. Prevalence, mean and median values should be in bold. Note that the Epi Info outputs are arranged in the way that results for men are displayed first, then women, then both sexes.</p> <ul data-bbox="555 645 1410 1153" style="list-style-type: none"><li data-bbox="555 645 1410 790">• For lines on the Fact Sheet displaying prevalence values: Fill in the "Row %" along with the confidence intervals ("LCL" and "UCL") from the Epi Info outputs for the total age groups for both sexes, men and women.</li><li data-bbox="555 790 1410 1014">• For lines on the Fact Sheet displaying mean values: Fill in the mean values along with the confidence intervals ("Confidence Limits", "Lower" and "Upper") from the Epi Info outputs for the total age groups for both sexes, men and women. Note that for outputs displaying mean values, the values for total age groups are displayed in separate tables.</li><li data-bbox="555 1014 1410 1153">• For the lines on the Fact Sheet displaying median values: Fill in the median values along with the 25<sup>th</sup> and 75<sup>th</sup> percentile ("25%" and "75%") from the Epi Info outputs for the total age groups for both sexes, men and women.</li></ul>

**Note:** The generic Epi Info programs AgeRange2564 (or AgeRange1564) and MissingAgeSex should be run prior to running any other programs (see Part 3, Section 6).

## Preparing and Distributing the Data Book

---

**Introduction** The Data Book is a full tabulation of all the results from the questions and measurements in the STEPS Instrument.

---

**Purpose** The purpose of the Data Book is to:

- compile a complete, standardized set of results relating to each question and measurement in the instrument;
- provide a first step in the reporting process from which results for the site report can be extracted.

---

**Intended audience** The Data Book should not be considered as a final document for presenting your STEPS survey results. It is simply an intermediate step and helps in preparing your site report. For a review of your Data Book, it is recommended that you share copies with:

- WHO Regional Office, STEPS focal point and the
- WHO Geneva STEPS team.

---

**Standardized format and results** The STEPS Data Book template is available in Part 6, Section 3D. Use this template to present the results of your STEPS survey in a standardized format.

If you have added optional questions that are not represented in this template, you should add additional tables displaying results for those.

---

**Data Book layout** The Data Book consists of:

- the title page;
- the second page including the table of contents and information on what Epi Info programs need to be run first;
- ten topics, one for each section of your questionnaire.

Under each topic, there are several tables that relate to the topic and the corresponding questions. The table headings include a brief description of what is contained in the table, followed by the instrument question(s) that are needed to prepare these results.

All tables in the Data Book include another brief heading in the first row, an indication on whether the table is for men, women, both sexes, or all three, a column with the age groups, a column with the "n" for the subgroups, and columns with the actual results (mean, median or prevalence) with confidence intervals. Note that the tables in the section Demographic Information Results don't contain confidence intervals.

---

*Continued on next page*

## Preparing and Distributing the Data Book, Continued

### Data Book layout (cont.)

The analysis information below each table displays the standard codes of the questions used and the Epi Info program names for the programs that need to be run to get the results for the corresponding table.

### Process

Generic programs in Epi Info have been written to generate all the tables for the Data Book. The table below lists the process required to fill in the tables in the Data Book.

Stage	Description
1	Insert the name of your site/country in the title page of the Data Book.
2	Identify which tables can be completed in the Data Book by referring to the "Analysis Information" section below each data table and using your site-specific instrument.
3	Run the program for each table identified in step 1 above, and save the output (for information on how to run Epi Info programs see Part 3, Section 6).
4	<p>Copy the output of each program into the corresponding table in the Data Book. Note that the Epi Info outputs are arranged in the way that results for men are displayed first, then women, then both sexes.</p> <ul style="list-style-type: none"> <li>• Fill in the "n" for all age groups. The "n" indicates the number of individuals that have responded to a specific question; it is the denominator for the prevalence/mean/median values for each age group. In Epi Info outputs displaying prevalence values, you can find the "n" in the "TOTAL" column for each age group, in the outputs displaying mean and median values, you can find the "n" in the "Obs" or "Count" column.</li> <li>• For tables displaying prevalence values: Fill in the "Row %" along with the confidence intervals ("LCL" and "UCL") from the Epi Info outputs for the corresponding age groups.</li> <li>• For tables displaying mean values: Fill in the mean values along with the confidence intervals ("Confidence Limits", "Lower" and "Upper") from the Epi Info outputs for the corresponding age groups. Note that for outputs displaying mean values, those values for total age groups are displayed in a separate table.</li> <li>• For the tables displaying median values: Fill in the median values along with the 25<sup>th</sup> and 75<sup>th</sup> percentile ("25%" and "75%") from the Epi Info outputs for the corresponding age groups.</li> </ul>

**Note:** The generic Epi Info programs AgeRange2564 (or AgeRange1564) and MissingAgeSex should be run prior to running any other programs (see Part 3, Section 6).

For assistance in generating code to analyse site-specific questions, contact the STEPS team.

## Preparing and Distributing the Site Report

---

**Introduction** The site report is the main comprehensive report for the whole STEPS chronic disease risk factor survey and must be produced at the end of the STEPS survey.

A template that helps preparing the STEPS site report is in Part 6, Section 3E.

---

**Purpose** Use the site report to present the following information:

- the overall rationale
- scope of the survey
- the sampling design used
- detailed methods of data collection
- detailed results of the survey
- implications for future health and planning
- appendices including the site-specific instrument.

---

**Intended audience** It is recommended that you distribute the site report widely. Consider sending copies to:

- relevant government bodies and sponsoring organizations;
- agencies and organizations that are likely to use the information to promote chronic disease prevention and control;
- public, governmental and institutional (university) libraries;
- press and other media (newspapers, radio and television);
- websites of any sponsoring bodies;
- WHO STEPS Regional Office and the WHO Geneva STEPS team.

---

**Content guide** The table below lists each of the main parts that should be included in the site report.

- cover and content pages
- executive summary
- background
- rationale of the survey
- scope, sampling methods and implementation
- analytical methods
- results
- discussion
- conclusions and recommendations
- references
- appendices

**Note:** Guidelines for completing each of these parts are provided in the STEPS site report template in Part 6, Section 3E.

---

## Progress Report

---

**Introduction** A progress report may be prepared, but only where interviewing takes longer than the usual eight week time period. The report should explain the reason for delays and provide an update on participation rates and data collection quality.

---

**Intended audience** Progress reports are intended for the STEPS team and STEPS Coordinating Committee only, although details may be used to inform stakeholders of progress.

---

**Content guide** Progress reports should be prepared in conjunction with data collection teams and the data entry supervisor.

Follow the guidelines in the table below to help prepare content for a progress report.

<b>Heading</b>	<b>Guidelines for completion</b>
Tables and plots	Describe in tables and/or plots: <ul style="list-style-type: none"><li>• locations where interviewing is:<ul style="list-style-type: none"><li>– complete</li><li>– ongoing</li><li>– awaited;</li></ul></li><li>• number of people selected in the sample to date;</li><li>• number and proportion who have completed;</li><li>• number who have refused or are not able to be contacted (non-response);</li><li>• number for whom no attempt has yet been made to contact.</li></ul>
Successes and problems	Identify successes and problems with: <ul style="list-style-type: none"><li>• data collection</li><li>• data entry</li><li>• data analysis.</li></ul>

---



## Section 5: Archiving

### Archiving your STEPS Materials

---

**Introduction** Once the survey is completed and before the team is disbanded, all records need to be properly stored in order to prevent loss.

---

**Policies and systems** Most governments and large organisations will have their own established archival systems, in which case their facilities are likely to be your best long-term storage option. Investigate storing your data at:

- Ministry of Health
  - WHO country office
  - WHO regional office.
- 

**Archival period** Decide on the archival period. The duration may have been specified by your ethics authority. If not, consider twelve years.

This is long enough for data to be available for further STEPS surveys, and long enough to investigate query from the results.

---

**Checklist** Use the checklist below to help ensure all necessary steps have been completed.

Step	Action	✓
1	Decide on the duration of storage.	
2	Box up all: <ul style="list-style-type: none"><li>• instruments</li><li>• manuals</li><li>• interviewing materials</li><li>• printed versions of all files.</li></ul>	
3	Label all the boxes clearly with: <ul style="list-style-type: none"><li>• name and date of the project;</li><li>• box contents;</li><li>• names and contact details of site coordinator and one other member of the coordinating committee.</li></ul>	
5	Determine who is entitled to have access to the archive.	
6	Place a copy of the form to apply for access in each box.	
7	Provide copies of electronic files (without personal identifiers) to WHO Geneva STEPS team.	
3	Inform all interested parties where the information is stored.	

**Note:** Make sure that participant identification information is never stored in the same location (electronically and in paper form) with the rest of the dataset.

---

