

## Section 6: Data Analysis Guide

### Overview

---

**Introduction** This section provides general guidelines for the data analyst as well as a basic introduction to Epi Info. For more specific instructions on how to proceed with the analysis of your STEPS data, see Part 4, Section 2.

---

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

- data analyst
- STEPS Site Coordinator
- statistical adviser.

---

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

Topic	See Page
General Information	3-6-3
Introduction to Epi Info	3-6-4

---

## Overview

---

### Introduction

The data analyst is responsible for :

- creating the database
  - cleaning and weighting the data
  - producing the completed fact sheet and data book.
- 

### What you will learn

In this course you will learn how to setup and use Epi Info to analyse your STEPS survey data.

---

### Learning outcomes

The learning outcomes of this section are to be able to:

- navigate the Epi Info Analysis module
- run basic commands in Epi Info Analysis
- run the generic analysis programs provided by the Geneva STEPS team.

The generic analysis programs perform basic calculations needed to complete the Fact Sheet, Data Book, and site report.

---

### Other data analyst materials

This guide is to be used in conjunction with the following sections in the STEPS surveillance manual. These sections provide full background detail and instructional material on the following topics.

<b>Topic</b>	<b>Part, Section</b>
Preparing the Sample	Part 2, Section 2
Preparing the Data Analysis Environment	Part 2, Section 5
Data Analysis	Part 4, Section 3
Fact Sheet Analysis Guide	Part 6, Section 3B
Data Book Template	Part 6, Section 3D

---

## General Information

---

**Introduction** It is important that the data analyst has some background information on the STEPS survey as this may impact the way they analyse the data. The general information that the analyst needs and where they can find this information is described below.

---

**Scope of survey** The scope of the survey should be available in the implementation plan. The STEPS Site Coordinator will also have this information. The data analyst must understand the scope of the survey so that the results of the analysis reflect the scope.

---

**Sample method** It is essential that the data analyst understands what sampling method was used for the STEPS survey so that the data can be properly weighted. The analyst should be familiar with the Interview Tracking Form Excel workbook (interview\_tracking\_form.xls) and the STEPS sampling Excel workbook (STEPSsampling.xls).

The sampling information should already be documented in the STEPS Implementation Plan and/or supporting documents and available for the analyst. If it is not documented, consult the STEPS Site Coordinator and make sure the information is documented right away. It is critical information and needs to be documented.

---

**Assisting with fact sheet and site report** The data analyst should also assist the Site Coordinator with the Fact Sheet and site report. Liaise with the Site Coordinator to identify the data analyst's roles and responsibilities, see Part 4, Section 4 for more information.

---

## Introduction to Epi Info

---

**Introduction** For the analysis of data, the STEPS team recommends and supports Epi Info, a purpose-built, free, public-domain software package. While Epi Info has a broader range of functions, this manual will only explain how to perform data analysis in Epi Info using the Analysis module in Epi Info.

---

**Rationale** The decision for choosing Epi Info was made in light of its advantages, some of which are listed below.

- Windows-based
  - recent release of Epi Info, supported by developers
  - has data analysis capability in line with STEPS requirements
  - can appropriately adjust for complex sampling designs.
- 

**Topics covered** The following topics are covered in this tour of Epi Info:

- basic terminology
  - opening the Analysis module and Analysis module screen components
  - software settings and basic commands
  - creating a new or derived variable
  - displaying a variable
  - obtaining basic descriptive statistics on a variable
  - recoding a variable
  - displaying data in a graph
  - running saved programs
  - selecting a subset of records in a dataset
  - saving and printing outputs.
- 

**Downloading & Installing Epi Info** For instructions on downloading and installing Epi Info, please see Part 2, Section 5, Preparing the Data Analysis Environment.

---

*Continued on next page*

## Introduction to Epi Info, Continued

### Terminology

Some of the specific Epi Info terms used are described in the table below.

<b>Term</b>	<b>Description</b>
Command	Predefined term in Epi Info syntax (language) that tells Epi Info how to manipulate or analyse your data (e.g. LIST, SELECT).
Program (.pgm)	Syntax files that can be saved in a separate text file or within an Access database. Contain a series of commands in Epi Info syntax to manipulate and analyse data.
Project	The name of the actual Access database (.mdb) file. All the data and related programs are stored within the project.
Variable	Any characteristic or attribute that can be measured. For STEPS datasets, most variables correspond to one question on the Instrument.

### To open Epi Info

To open the Epi Info Analysis module double click on the Epi Info icon on your desktop and click the "Analyze Data" button in the lower left-hand section of the screen.

Alternatively, you can open the Analysis module directly by navigating to your Epi Info program folder (e.g. C:\Epi\_Info) and double-clicking on the "analysis.exe" file.

### Screen Components

The Analysis module of Epi Info has three main components divided into the following three windows:

<b>Window</b>	<b>Function</b>
Analysis	Contains all the commands that can be used during analysis.
Analysis Output	Displays the results of a program once it has been run.
Program Editor	Displays the code of saved programs and can be used to write new programs.

*Continued on next page*

## Introduction to Epi Info, Continued

---

### Software Settings

Follow the steps below to set Epi Info to exclude missing data and to provide the appropriate output for weighted analyses.

Step	Action
1	In the Analysis window, click on Analysis Commands>Options>Set. It is located at the very bottom the Analysis window.
2	In the SET window, set "Yes as" to "Yes", "No as" to "No", and "Missing as" to "Missing".
3	Ensure all 6 check boxes immediately beneath the "Yes as" drop-down box are checked.
4	Set "Statistics" to "Advanced".
5	Ensure the check box for "Include missing" is NOT checked.
6	Set "Process records" to "Normal (undeleted)".
7	Click "Save all".

---

### Open a dataset

Follow the steps below to open a dataset that is stored as a data table within an Access database.

Step	Action
1	In the Analysis window, click on Analysis Commands>Data>Read (Import) to open the READ window.
2	Set the "Data format" to "Epi 2000".
3	Click on "Change Project" and find and select your Access .mdb file (e.g. STEPS.mdb).
4	Click on the name of your dataset (e.g. MasterDataSet) from those listed.
5	Click "OK".

**Note:** The file path, number of records and date/time will be displayed in the Analysis Output window once the dataset is opened.

---

*Continued on next page*

## Introduction to Epi Info, Continued

### Create a new or derived variable

Follow the steps below to create and assign values to a new or derived variable (e.g. BMI).

Step	Action
1	In the Analysis window, click on Analysis Commands>Variables>Define to open the DEFINE window.
2	In the DEFINE window, type the name of your new variable in the space provided (e.g. BMI).
3	Ensure "Scope" is set to Standard and click "OK".
4	In the Analysis window, click on Analysis Commands>Variables>Assign to open the ASSIGN window.
5	In the "Assign Variable" field, select your newly defined variable (e.g. BMI) from the drop-down list.
6	In the "Expression" field type the formula to compute the values of your new variable (e.g. weight/height*height). To use existing variables (e.g. weight and height) in your formula, select them from the "Available Variable" drop-down list.
7	Click "OK".

### List all variable values

Follow the steps below to list the value of a variable for all records in the analysis output.

Step	Action
1	In the Analysis window, click on Analysis Commands>Statistics>List to open the LIST window.
2	Choose the variable(s) you wish to list from the "Variables" drop-down list or click the check box "All (*) Except" to list all variables.
3	Click "OK". A list of the chosen variables will be displayed in the Analysis Output window.

### Create a frequency table for a variable

Follow the steps below to create a frequency table for a variable containing a list of all values for a given variable and the frequency of each value.

Step	Action
1	In the Analysis window, click on Analysis Commands>Statistics>Frequencies to open the FREQ window.
2	Choose the variable(s) you wish to list from the "Frequency of" drop-down list (select * to list all variables).
3	Click "OK". A frequency table(s) will be displayed in the Analysis Output window.

*Continued on next page*

## Introduction to Epi Info, Continued

### Create a means analysis for a variable

Follow the steps below to perform a means analysis on a variable. The output of this command provides a frequency table as well as the mean, variance, standard deviation, and quartile values for the variable.

Step	Action
1	In the Analysis window, click on Analysis Commands>Statistics>Means to open the MEANS window.
2	Choose the variable(s) you wish to list from the "Means of" drop-down list (select * to list all variables).
3	Click "OK". A frequency table(s) and means analysis will be displayed in the Analysis Output window.

### Recode a variable

Follow the steps below to recode a variable (e.g. age to Agerange).

Step	Action
1	In the Analysis window, click on Analysis Commands>Variables>Recode to open the RECODE window.
2	Select the variable <b>from</b> which you want to recode (e.g. age) from the "From" drop-down list.
3	Select the variable <b>to</b> which you want to recode (e.g. Agerange) from the "To" drop-down list.
4	Complete one line in the table for each value or range of values for your "From" variable that you wish to recode. It is recommended that all possible values for the "From" variable are assigned a corresponding value in the "To" variable to avoid missing values in the "To" variable.  <b>Example:</b> To recode age values of 25 to 34 to the Agerange value of "25-34", type 25 for "Value", type 34 for "To Value", and type 25-34 for "Recode Value". To create further values for Agerange (e.g. 35-44, 45-54), complete additional rows in the table as needed.
5	Click "OK" when finished.

### Graph variables

Follow the steps below to graph variables.

Step	Action
1	In the Analysis window, click on Analysis Commands>Statistics>Graph to open the GRAPH window.
2	Select "Graph type" from the drop-down list (e.g. bar for binary or points to depict continuous variables).
3	In the "X axis" section, select the X axis variable from the "Main variable(s)" drop-down list.
4	In the "Y axis" section, set "show value of" to "Count".
5	Fill in labels and titles if desired and click "OK".

*Continued on next page*

## Introduction to Epi Info, Continued

**Run programs** Follow the steps below to run a saved program.

Step	Action			
1	In the Analysis window, click on Analysis Commands>User-Defined Commands>Run Saved Program to open the RUNPGM window.			
2	Your program can either be stored in a program file (.pgm) or within an Access database file (.mdb). The table below shows how to run a program depending on the type of file in which it is saved.			
	<b>Program (.pgm) file</b>		<b>Access Database</b>	
	<b>Step</b>	<b>Action</b>	<b>Step</b>	<b>Action</b>
	2.1a	Click on the grey box to the right of the "Filename" field to search for your .pgm file.	2.1b	Click on the grey box to the right of the "Filename" field to search for your .mdb file.
	2.2a	Set "Files of type" to .pgm.	2.2b	Set "Files of type" to .mdb.
	2.3a	Once you have found your file, click "OK" in the RUNPGM window.	2.3b	Once you have found the database, select the program from the "Program" drop-down list and click "OK".

**Select a subset of a dataset** Follow the steps below to select a subset of a dataset. The SELECT command will stay in effect until another SELECT command is called.

Step	Action
1	In the Analysis window, click on Analysis Commands>Select/If >Select to open the SELECT window.
2	Complete the "Select Criteria" field with the desired equation (e.g. C1=1), use the drop-down list in the "Available Variables" field to select variables in your dataset for your equation.
3	When your equation is complete, click "OK".
4	To cancel the SELECT command, open the SELECT window again and click "OK", leaving the "Select Criteria" field blank.

*Continued on next page*

## Introduction to Epi Info, Continued

---

**Save a dataset** Follow the steps below to save a dataset after you have opened it and modified it.

Step	Action
1	In the Analysis window, click on Analysis Commands>Data>Write (Export) to open the WRITE window.
2	Set "Output Mode" to "Replace".
3	Click on the grey box next to the "File Name" field to search and select your data file (e.g. STEPS.mdb).
4	Select the name of the data table to which you would like to write in the drop-down list (e.g. MasterDataSet) or write the name of a new table in the "Data Table" field.

---

**Save outputs** Follow the steps below to save the output in one file.

Step	Action
1	In the Analysis window, click on Analysis Commands>Output>RouteOut to open the ROUTEOUT window.
2	Define or browse for an output filename where your analysis is to be stored and click "OK".

---

**Print outputs** Follow the steps below to print outputs.

Step	Action
1	In the Analysis window, click on Analysis Commands>Output>PrintOut to open the RPRINTOUT window.
2	Click "OK". The results will be printed.

---