Section 2: Data Analysis

Overview

Introduction   The WHO STEPS team provides a suite of data analysis tools to assist you with the analysis of your STEPS data. While using these tools is not required, it is strongly recommended to use them not only to expedite the data analysis process but, more importantly, to ensure that the descriptive analysis is performed in a standardized way. Countries looking to develop their own data analysis tools or perform more complex analyses, are encouraged to use the STEPS tools to perform the standard descriptive analysis first.

This section provides a detailed overview of the data analysis tools available from the WHO STEPS team.

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

- data analyst
- statistical adviser
- STEPS Survey 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 .

Data analysis software   It is recommended to use Epi Info for data analysis (version 3.3 or higher), supplemented by Microsoft Access. (Note that currently Epi Info 7 does not support Microsoft Access project files containing analysis programs and therefore cannot yet be used with the STEPS analysis tools.)

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.

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

Overview, Continued

Tasks and timeframes

Data analysis cannot begin until the data has been cleaned and weighted (see Part 4, Section 1). Once that task it complete, the data analysis can be completed with the standard STEPS tools within the span of a week. Additional time would be needed if more complex analyses are going to be done. Note this does not include time to write the survey report (see Part 4, Section 3).

In this section

This section covers the following topics:

Topic	See Page
Preparing the Dataset	4-2-3
Epi Info Overview	4-2-5
Completing the Fact Sheet	4-2-8
Completing the Data Book	4-2-9

Part 4: Data Management and Analysis, Reporting and Disseminating Results
Section 2: Data Analysis
Last Updated: 26 January 2017

4-2-2
WHO STEPS Surveillance

Preparing the Dataset

Introduction	Once the dataset has been cleaned and weighted (see Part 4, Section 1), it must be prepared for analysis in Epi Info in order to use the standard STEPS analysis tools.
Epi Info Analysis Programs	The standardized analysis programs provided by the STEPS team are encapsulated in an Access file containing a table of all the Epi Info analysis programs. This Access file is available on the STEPS website.
Preparing files and folders	Once you have downloaded the Epi Info analysis programs, you must rename the file "STEPS.mdb". This file must be located in the following location on your computer in order for the analysis programs to work: C:\STEPS\EpiInfo Additionally, you should create a folder called "Output Tables" and place it in this folder as well. This is where Epi Info will save all of your analysis output when you perform your analysis in Epi Info using the provided programs.
Create mandatory variables	In order to use the Epi Info analysis programs the following variables must be in your dataset. These variables are included in most, if not all, of the data analysis programs and therefore the programs will not be able to run if they are missing. For the last 5 variables in the list, please refer to Part 4, Section 1 for more information about the development of these variables.

Variable	Description
Age	Age of the participant, calculated from date of interview and date of birth (if available, otherwise taken from the question on age of the participant).
Agerange	The age range into which the participant falls (e.g. "18-29", "30-44"). This is a text variable.
Sex	A text variable containing the values "Men" and "Women", generated from the variable C1 (sex) in the dataset.
Valid	A flag variable used to indicate which records should be included in all analyses. Records with either age or sex missing or age out of the age range of the survey receive a value of 2. Records with neither age or sex missing and age within the age range of the survey receive a value of 1.
WStep1	The complete analysis weight for Step 1 variables.
WStep2	The complete analysis weight for Step 2 variables.

Continued on next page

Part 4: Data Management and Analysis, Reporting and Disseminating Results   4-2-3
Section 2: Data Analysis   WHO STEPS Surveillance
Last Updated: 26 January 2017

Preparing the Dataset, Continued

Create mandatory variables (cont.)

Variable	Description
WStep3	The complete analysis weight for Step 3 variables.
PSU	Unique identifiers for sampling units above the household level (e.g. villages or enumeration areas).
Stratum	Unique identifiers for sampling units above PSU (e.g. districts, urban/rural). If not needed, simply set Stratum=1 for all records.

Note that there are simple Epi Info programs included within the provided analysis programs to generate the first 4 variables in the above table, if you wish to create these within Epi Info. You will need to run **AgeRange 1869** and then **MissingAgeSex**. The next topic in this Section provides instructions on how to run programs in Epi Info.

Import data

Once the necessary folders have been created and the STEPS.mdb file has been placed in the correct location on your computer, you will need to import your data file into the STEPS.mdb file.

Prior to importing the file, count the number of variables (columns) in your dataset to confirm the number is not greater than 255 (the maximum allowable in Microsoft Access). If the number of variables is greater than 255, you will need to split your dataset into two data tables, each with Participant ID. Please contact the WHO Geneva STEPS team for help with splitting your dataset and making the necessary modifications to the analysis programs.

To import your data from Excel to Access, follow the instructions in the table below.

Step	Action
1	Open the STEPS.mdb file and go to the External Data tab. Click on "Excel" under this tab.
2	In the pop-up window, locate your Excel file and tick the option "Import the source data into a new table in the current database." Click "OK".
3	Click "Next" on the first screen of the Import Spreadsheet Wizard pop-up window.
4	Tick the option "First Row Contains Column Headings" and click "Next".
5	Click "Next" on the next screen of the pop-up window. Then pick the option "No primary key" and click "Next".
6	Enter "MasterDataSet" in the Import to Table field. Click "Finish".

Contact the WHO STEPS team if your dataset is not already in Excel and you need assistance converting it to Excel in order to import it into Access.

Part 4: Data Management and Analysis, Reporting and Disseminating Results
Section 2: Data Analysis
Last Updated: 26 January 2017

4-2-4
WHO STEPS Surveillance

Epi Info Overview

Introduction   Epi Info is a free software package developed by the US Centers for Disease Control (CDC). While Epi Info has a broad range of functions, we are only referring to the Analysis module within this guide. Epi Info was chosen over a decade ago as the statistical software for which the STEPS tools were designed given that it is free, easy to use, and can appropriately adjust for complex sample designs.

This overview refers to Epi Info 3.5, which is freely available on US CDC website and on the STEPS website. (Note that currently Epi Info 7 does not support Microsoft Access project files containing analysis programs and therefore cannot yet be used with the STEPS analysis tools.)

Additional Resources   This overview will only cover the basic skills needed to run the analysis programs and locate and interpret the results. There is a more in-depth Epi Info Training Guide available on the STEPS website.

Analysis settings   Prior to running any analysis programs, you will need to make a slight change to the settings of the Epi Info Analysis module in order to have confidence intervals appear with your weighted analyses.

In the Epi Info Analysis module, click on "Set" at the bottom of the list of commands down the left-hand side of the screen. Change the Statistics option to "Advanced" and then click "OK" to save the change. You should only have to make this change once on your machine.

Running Analysis Programs   Follow the instructions in the table below to run an analysis program located within your STEPS.mdb file.

Step	Action
1	Open the Epi Info Analysis module.
2	Click on "Open" in the Program Editor at the bottom of the screen.
3	In the Read Program pop-up window, you must check that the Project File is your STEPS.mdb file. If it is not, click on "Change Project" and locate and select your STEPS.mdb file. You should only have to make this change once on your computer until you work with another data file.
4	Select the program you wish to run from the Program drop down list and click "OK".
5	You will now see the program code in the Program Editor. Click "Run" to run the program. It should complete in a few seconds.
6	The output of the program will have the same name as the program and be located in the C:\STEPS\EpiInfo\Output Tables folder.

Continued on next page

Part 4: Data Management and Analysis, Reporting and Disseminating Results
Section 2: Data Analysis
Last Updated: 26 January 2017

4-2-5
WHO STEPS Surveillance

Epi Info Overview, Continued

Reading Epi Info Output

Nearly all output from the standard analysis programs will contain tables with results for men, then women, then both sexes. Some analyses produce multiple tables for each of these three groups.

An example prevalence table is shown below:

Men: Smoking status			
Forward			
Next			
agerange	**D**		
	1) daily and non-daily smokers	2) non-smoker	TOTAL
18-44	52	388	440
Row %	12.553	87.447	100.000
Col %	64.666	81.198	78.673
SE %	2.260	2.260	
LCL %	8.084	82.977	
UCL %	17.023	91.916	
Design Effect	2.042	2.042	
45-69	47	145	192
Row %	25.304	74.696	100.000
Col %	35.334	18.802	21.327
SE %	4.549	4.549	
LCL %	16.306	65.697	
UCL %	34.303	83.694	
Design Effect	2.091	2.091	
TOTAL	99	533	632
Row %	15.273	84.727	100.000
Col %	100.000	100.000	100.000
SE %	1.963	1.963	
LCL %	11.390	80.845	
UCL %	19.155	88.610	
Design Effect	1.878	1.878	

The **Total** column provides the total number of respondents included in the analysis for each age group as well as overall. In the above example, 632 men were included in the analysis, 440 men aged 18-44 and 192 age 45-69.

Row % provides the percentage of respondents in each category (column) for the age group or overall. In the above example, 12.6% of men aged 18-44 were current smokers (daily or non-daily smokers). Row % sums across the row. We generally use Row % in the STEPS reporting documents.

Col % provides the percentage of respondents within each category that fall into each age group. In the above example 64.7% of current smokers are aged 18-44. We do not generally use Col % in the STEPS reporting documents.

Continued on next page

Part 4: Data Management and Analysis, Reporting and Disseminating Results
Section 2: Data Analysis
Last Updated: 26 January 2017

4-2-6
WHO STEPS Surveillance

Epi Info Overview, Continued

Reading Epi Info Output (cont.)

95% confidence intervals are provided in the **LCL%** and **UCL%** rows of the output tables. In the above example, the 95% confidence interval for the 12.6% of men aged 18-44 who currently smoke is 8.1-17.0.

Compacting the dataset

After you have run several analysis programs, your STEPS.mdb file will expand in size and become very large. When you have finished running your analysis programs (or if there is a noticeable slow down while you are in the middle of your analysis programs), close Epi Info and open your STEPS.mdb file in Access.

To compact your data file back to its usual size, go to the File tab and then click on "Compact & Repair". After a few seconds your file should be back to its usual size.

Completing the Fact Sheet

Introduction   The STEPS Fact Sheet is a short summary of key results of the STEPS survey. It is intended to be distributed widely to draw attention to the survey results and highlight issues that will be covered in more depth in the survey report.

The STEPS Fact Sheet template and the STEPS Fact Sheet analysis guide are located in Part 6, Section 3 of the Manual.

Fact Sheet layout   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. Additionally, at the very end of the Fact Sheet there are contact details of the country STEPS Survey Coordinator. Be sure to complete these pieces of text as well as the data fields.

The main body of the Fact Sheet contains a small number of indicators for each of the behavioural and metabolic risk factors covered in the survey. Additionally, there are a few indicators in which behavioural and metabolic risk factors are combined. For each indicator, results are presented for the overall sample, for all males and for all females.

Fact Sheet Analysis Guide   The Fact Sheet Analysis Guide has been developed to assist data analysts in 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 question(s) required to calculate that particular indicator, and one column that includes the names of the Epi Info program that needs to be run to produce results for that indicator.

Procedure   To complete the STEPS Fact Sheet, it is recommended to first review the Fact Sheet Analysis Guide to determine which Epi Info programs needs to be run. If the relevant questions have been dropped from the local STEPS Instrument, the related line in the Fact Sheet can be removed.

Once all Epi Info analysis programs have been identified, follow the instructions provided in the Epi Info Overview earlier in this Section to run each program.

Point estimates (prevalences or means) should be rounded to one decimal point and be presented with 95% confidence intervals (also rounded to one decimal point).

Completing the Data Book

Introduction   The Data Book is a full tabulation, by age range and sex, of the data from all the questions and measurements in the STEPS Instrument. It is intended to serve as the basis for the country report, to guide the writers on what results to include and highlight in the report. While selected tables may be included in the body of the country report, it can be included in its entirety as an appendix to the report.

The template of the STEPS Data Book is located in Part 6, Section 3 of the Manual. Additional data book pages for the optional modules are available from the WHO STEPS team or can be downloaded from the STEPS website.

Data Book layout   After the title page and table of contents, the remaining pages of data book contain for each table of results a title, description of the table (including the full question text from the questionnaire) and analysis information for the data analyst. Results are presented in the same order as the questions in the questionnaire.

The analysis information contains the standard question codes for the questions required for the given analysis as well as the name of the analysis program that needs to be run to complete the table. The analysis information can be deleted once the data book has been completed.

Each of the data tables contains results for each age group for both sexes and for each sex separately. For each age-sex group in the table, the point estimate is given (prevalence or mean) along with 95% confidence interval (except for the demographic information) and the "n" (the total number of individuals included in the analysis for that age-sex group, for example, the total number of men aged 18-29 included in the analysis).

Procedure   To complete the STEPS Data Book, it is recommended to first review document to determine which Epi Info programs needs to be run. If the relevant questions have been dropped from the local STEPS Instrument, the related table(s) can be removed. Additionally, if after running an analysis program the "n" is very small (i.e. less than 50 respondents), you can either delete the entire table or, if there are enough respondents, only show results for the overall age group.

Once all Epi Info analysis programs have been identified, follow the instructions provided in the Epi Info Overview earlier in this Section to run each program.

Both the point estimates (prevalences or means) and the 95% confidence intervals should be rounded to one decimal point.

Part 4: Data Management and Analysis, Reporting and Disseminating Results
Section 2: Data Analysis
Last Updated: 26 January 2017

4-2-9
WHO STEPS Surveillance

Completing the Data Book, Continued

Formatting macros

There are formatting macros available from the WHO Geneva STEPS team to assist you in putting together your Data Book. These macros are located in two separate Excel files, one containing macros to format unweighted tables, the other containing macros to format weighted tables. Be sure only one of these is open on your computer at a time, else the macros may not work correctly.

To format a table from the Epi Info output, follow the instructions in the table below:

Step	Action
1	Copy the entire table from Epi Info and paste it into cell A1 in the PASTE sheet of the Excel macro file.
2	Run the macro that corresponds to your particular table. Refer to the Instructions sheet in the Excel file to see which macro should be run. Which macro to run depends on the type and size of the output table. For example, to format a means output table, press ctrl + m while on the PASTE sheet to run the macro.
3	The macro should take a few seconds to run at most. Once it is completed, you will have a formatted table you can now copy and paste directly into the corresponding table in the Data Book. Be sure to highlight all the relevant squares in the Data Book table before pasting.

Part 4: Data Management and Analysis, Reporting and Disseminating Results   4-2-10
Section 2: Data Analysis   WHO STEPS Surveillance
Last Updated: 26 January 2017