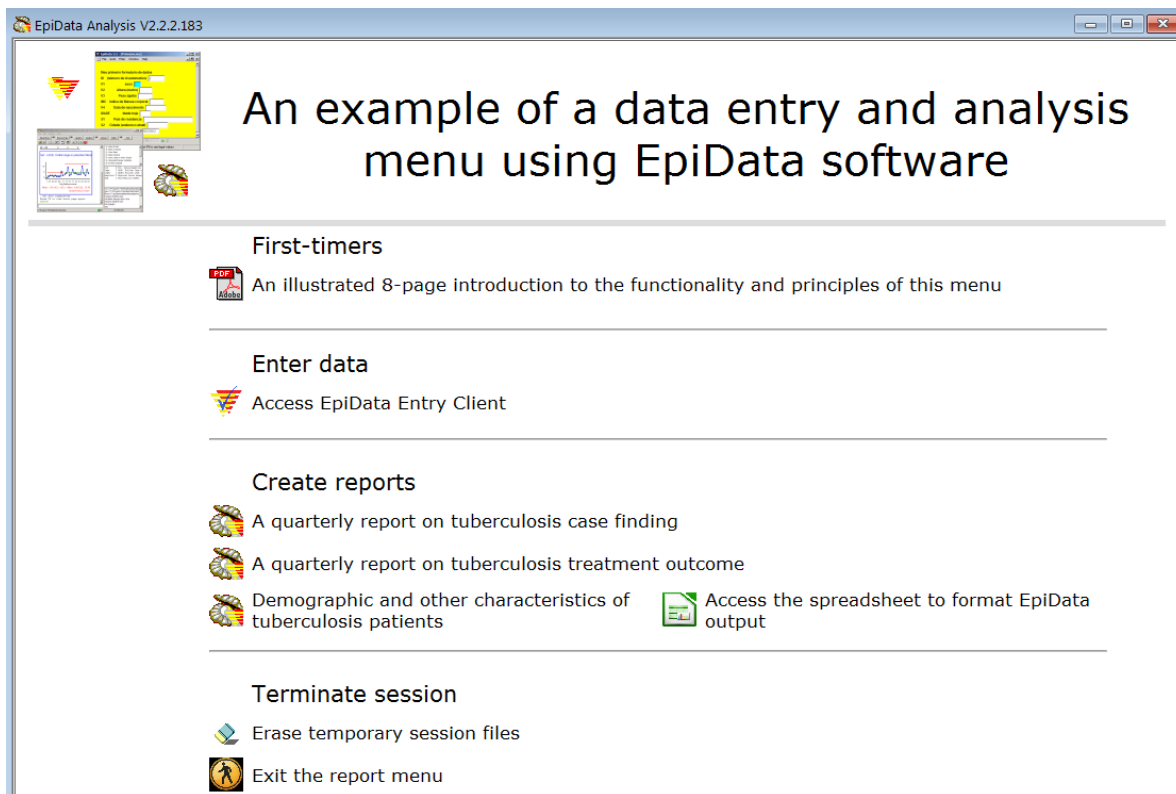


A example of a data capture and reporting system menu using EpiData software

Hans L Rieder



Introduction

This package uses the capability of EpiData Analysis to create an HTML-based menu allowing functionality from within a browser using hyperlinks embedded into icons. If clicked, something is executed. The menu allows capturing data recorded in the standard tuberculosis case register in an electronic data collection instrument made with EpiData Manager (available freely at <http://www.epidata.dk>). The entry form is accessed through the EpiData Entry Client. At any point in time, selections can be made to produce any number of customized reports from the captured data, by clicking on an icon using EpiData Analysis.

Enter data

Access EpiData Entry Client

If you click on the EpiData Entry Client icon (the two overlapping triangles with a check mark), the data collection instrument for the Tuberculosis Register opens (only the first few lines shown):

Data Entry form for the year 2006

Unique patient identifier

Treatment unit

Patient identifying serial number

Patient's age (yrs)

Patient's sex

Year of registration

Month of registration

Day of registration

Value Labels

Value	Label
1	Unit A01
2	Unit A02
3	Unit A03
4	Unit A04
5	Unit A05
6	Unit A06
7	Unit A07
8	Unit A08

The choice of colors for the fields is made in “Preferences” and is individual for the computer. Here, blue was chosen for the field which has the focus. Fields which cannot be entered (example here “Unique patient identifier”) [or fields for which the user is free to enter a value or not] (not applicable here) are colorless transparent, fields that require a value are colored reddish (example here “Patient identifying serial number”). Fields of the latter type are called “Must enter” fields in EpiData.

Hence, when opening the focus is on the first “Must enter” field which here is for the treatment unit which opens a drop-down menu. Note that in EpiData Entry Client (in contrast to the old version EpiData Entry 3.1) you can choose any sub-string from either the “Value” or the “Label”.

After the value for the first field has been entered, the cursor moves to the field “Patient identifying serial number” (pressing the “Enter” key is required if the allowable field length has not been used up). To increase efficiency, numeric coding is used whenever possible. The latter is such a field for a categorical variable. To remain unambiguous, a drop-down window pops up as soon as the cursor is in the field:

This is an integer field, but in the design in EpiData Manager it has been set to display “leading zeros”, i.e. after entry of the serial number 3 we get:

In the next field AGE we get a pop-up Note which is useful for continuous variables to let the entry person know what to enter for missing data (or here also what to do should the value exceed the field length (which was set here to 2, thus the entry person needs to know what to enter if persons are aged 100 years or older).

After the value for the field “Year of registration” is entered, there is sufficient information to create the unique identifier (top variable):

Note that as a “No enter” field the value is grayed out.

If the data entry person tries to bypass a “Must enter” field without entering a value, the field turns yellow:

Year of registration

Month of registration

Day of registration

For continuous variables, a warning “Field cannot be empty” pops up.

A “Must enter” field can, however, be bypassed with the mouse, but in contrast to EpiData Entry 3.1 this will catch up with the entry person at the very end of data entry for the record. At this point, EpiData will check for completion of all “Must enter” fields, and if any one misses an entry, the cursor returns to the field without allowing saving the record:

Smear result at 5 months Negative

Smear result at 7 months or later

Principal treatment outcome Cured

Year of treatment outcome

Month of treatment outcome

Day of treatment outcome

Value	Label
0	Negative
0.1	Scanty, 1 AFB per 100 fields
0.2	Scanty, 2 AFB per 100 fields
0.3	Scanty, 3 AFB per 100 fields
0.4	Scanty, 4 AFB per 100 fields
0.5	Scanty, 5 AFB per 100 fields

Once all fields are completed, the entry person is prompted to save to disk:

Smear result at 7 months or later Negative

Principal treatment outcome Cured

Year of treatment outcome

Month of treatment outcome

Day of treatment outcome

Warning

Current record is modified.
Save record?

Yes No Cancel

If the user wishes to exit the module, there will be another prompt:

Data Entry form for the year 2006

Unique patient identifier

Treatment unit

Patient identifying serial number

Patient's age (yrs)

Patient's sex

Year of registration

Warning

Project data content modified.
Store project permanently on disk before exit?

Yes No Cancel

This is so because EpiData Entry Client writes the data during entry into memory and needs to write it now permanently to disk. This is thus different from EpiData Entry 3.1 where every record was definitely written to disk!

Create reports



A quarterly report on tuberculosis case finding



A quarterly report on tuberculosis treatment outcome



Demographic and other characteristics of tuberculosis patients



Access the spreadsheet to format EpiData output

Once you have entered some data, you wish to analyze them. To this end, we have in this example already a dataset from real observations over a two-year period that produces two standard quarterly report, one on case finding, the other on treatment outcome. You are prompted to enter the date components of the start and ending date for your analysis: the interval can be anything as long as it results in at least 2 records that are required to run the analysis. The output is both tabular and graphic.

A special type of analysis and display of output is provided with the third analysis which is a program / spreadsheet pair:



Demographic and other characteristics of tuberculosis patients



Access the spreadsheet to format EpiData output

Assume two variables, one on the specimen, and one on the result of microscopic examination:

Specimen type

Microscopy result

Name: spectype	Type: Integer
Label: Type of specimen	
Value	Label
1	Sputum
2	Bronchoalveolar lavage
3	Gastric lavage
4	Pus
5	Tissue
6	Urine
7	Stool
8	Other materials
9	Not recorded

Name: micro	Type: Integer
Label: Microscopy result	
Value	Label
0	Negative
1	1+ positive
2	2+ positive
3	3+ positive
4	Positive, not quantified
5	Scanty positive, not quantified
6	Scanty positive, quantified
9	Not recorded

If we make a cross-tabulation, what we get depends on the values that are available for both variables:

Type of specimen		
Microscopy result	Sputum	Total
1+ positive	2	2
2+ positive	2	2
Scanty positive, not quantified	1	1
Scanty positive, quantified	1	1
Total	6	6

Here we have 6 records, and for each the specimen type was sputum, 1 of 9 possible choices. Only 4 of the 8 possible microscopy results had been obtained for these 6 specimens. In other words, EpiData Analysis (and any other analysis software for that matter) displays only those row-column pairs for which there is a value for at least one of the two variables. An exhaustive display would show all possible columns and rows, including those with empty cells, as in this table of the same data:

Microscopy result	Type of specimen									Total
	Sputum	BAL	Gastric	Pus	Tissue	Urine	Stool	Other	Unknown	
Negative	0	0	0	0	0	0	0	0	0	0
1+ positive	2	0	0	0	0	0	0	0	0	2
2+ positive	2	0	0	0	0	0	0	0	0	2
3+ positive	0	0	0	0	0	0	0	0	0	0
Positive, not quantified	0	0	0	0	0	0	0	0	0	0
Scanty positive, not quantified	1	0	0	0	0	0	0	0	0	1
Scanty positive, quantified	1	0	0	0	0	0	0	0	0	1
Not recorded	0	0	0	0	0	0	0	0	0	0
Total	6	0	0	0	0	0	0	0	0	6

However, such a table cannot be produced with table commands. To get this type of display requires a work-around. Suffice to say here that the programs in the menu are designed to produce precisely an output containing all the necessary information to produce this extended type of a cross-table. In a first step, the EpiData Analysis program is run to generate the necessary output.

It is assumed that all data are entered into a single database (the preferred method). Therefore, the user must be enabled to choose the time frame for a report interactively. This report:



Demographic and other characteristics of tuberculosis patients



Access the spreadsheet to format EpiData output

gives a cross-tabulation of case status and several demographic characteristics. The corresponding menu line has two components: on the left is the access icon to run the EpiData Analysis program, and on the right the access icon to the spreadsheet to format the output generated in EpiData Analysis. Clicking the icon for EpiData Analysis (the open shell) starts running the program up to the first interactive prompt:

Enter beginning of registration period, day, month, year
Only a date between 1 Jan 2004 and 31 Dec 2005 is possible

Enter begin day of registration

OK Cancel

You are requested to enter date components with three successive prompts, first the day (shown here), then the month, and finally the year. Any legal combination (the illegal 31, 2, 2004 would result in an illegal date and break the flow of the program) is valid. After entering the three components, the start date is shown and the user is prompted to proceed in the same manner with the ending date:

Your chosen start date is 1 - 1 - 2004

Enter now your choice of the end of registration period, day, month, year
Only a date between 1 Jan 2004 and 31 Dec 2005 is possible

Enter end day of registration

OK Cancel

The source date for extraction is the date of registration. After selecting the period for analysis, the program runs its course and ends by popping up the data file abstraction in the browser:

	rowvarbxt	case	noncase
1	01_age_Q1	1014	627
2	01_age_Q2	1365	465
3	01_age_Q3	1367	395
4	01_age_Q4	1312	432
5	01_age_missing	0	2
6	02_sex_fem	1408	728
7	02_sex_male	3650	1192
8	02_sex_miss	0	1
9	03_site_pulm	5030	848
10	03_site_eptb	23	1072
11	03_site_miss	5	1
12	04_categ_new	4366	1491
13	04_categ_relapse	502	11
14	04_categ_failure	64	0
15	04_categ_default	22	2
16	04_categ_transfer	76	69
17	04_categ_other	8	48
18	04_categ_missing	20	300

together with a prompt to copy the content of the browser window to the clipboard:

Data compilation complete

Copy content of the browser window to clipboard and past into spreadsheet

F8: Go back to Menu

	rowvarbxt	case	noncase
1	01_age_Q1	1014	627
2	01_age_Q2	1365	465
3	01_age_Q3	1367	395

Clicking on the red circled intersection between columns and rows marks the entire file blue. Right-click to pick "Copy Selection to Clipboard":

	rowvar.txt	case	noncase
1			
2			
3			
4	01_age_04	1312	432
5	01_age_missing	0	2
6	02_sex_fem	1408	728
7	02_sex_male	3650	1192
8	02_sex_miss	0	1
9	03_site_pulm	5030	848
10	03_site_eptb	23	1072
11	03_site_miss	5	1
12	04_categ_new	4366	1491
13	04_categ_relapse	502	11
14	04_categ_failure	64	0
15	04_categ_default	22	2
16	04_categ_transfer	76	69
17	04_categ_other	8	48
18	04_categ_missing	20	300

Anything in the clipboard can be pasted into another place / file with the short-cut key combination **CTRL+v** (the “+” is only to show the combination for simultaneously holding down the “Ctrl” key and the letter “v”). Once copied, close the data browser window (clicking on “Close” or use the short-cut combination **ALT+c**). Finally, restore the menu with function key **F8**, getting you back to:



Demographic and other characteristics of tuberculosis patients



Access the spreadsheet to format EpiData output

The content of the clipboard contains all the necessary data, but the variable names are not that meaningful and the presentation is not nice. To get it nicely formatted (there is also functionality to a good formatted display) we use a spreadsheet. You are most likely familiar with the proprietary spreadsheet software Excel® from Microsoft. We don’t want to make ownership of proprietary software a requirement. It is also not possible to easily control the behavior of Excel® (we found it wanting during testing). To be free of restrictions, we thus placed a portable version of the free and open-source software suite LibreOffice© into the package. This frees the user providing unrestricted control and makes the package usable from any place (be it anywhere on a computer hard disk or an external USB drive), requiring no software installation: the package is ready to use. If you click the corresponding icon (circled in red):



Culture result and history



Access spreadsheet to paste and format analysis output

a preformatted OpenOffice Calc spreadsheet template opens (click only once, *be patient*, it is a “slow-opener”):

Note that this looks like the content from the EpiData data browser from where we copied the data, except that the cells are empty. The cursor is in the yellow colored cell A1. Read the content of the little box below (the first time at least) and then use **CTRL+v** to paste what you have in your clipboard and you get:

Text Import

Import

Character set: Unicode (UTF-16)

Language: Default - English (USA)

From row: 1

Separator Options

☒ Fixed width ☒ Separated by

☒ Tab ☐ Comma ☐ Semicolon ☐ Space ☐ Other

☐ Merge delimiters

Text delimiter: "

Other Options

☐ Quoted field as text ☒ Detect special numbers

Fields

Column type:

Standard	Standard	Standard
1 rowvartxt+case+noncase		
2 01_age_Q1+1014+627		
3 01_age_Q2+1365+465		
4 01_age_Q3+1367+395		
5 01_age_Q4+1312+432		
6 01_age_missing+0+2		
7 02_sex_fem+1408+728		
8 02_sex_male+3650+1192		
9 02_sex_miss+0+1		

Text Import

Import

Character set: Unicode (UTF-16)

Language: Default - English (USA)

From row: 1

Separator Options

☐ Fixed width ☒ Separated by

☒ Tab ☐ Comma ☐ Semicolon ☐ Space ☐ Other

☐ Merge delimiters

Text delimiter: "

Other Options

☐ Quoted field as text ☒ Detect special numbers

Fields

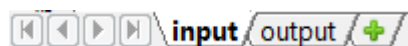
Column type:

Standard	Standard	Standard
1 rowvartxt	case	noncase
2 01_age_Q1	1014	627
3 01_age_Q2	1365	465
4 01_age_Q3	1367	395
5 01_age_Q4	1312	432
6 01_age_missing	0	2
7 02_sex_fem	1408	728
8 02_sex_male	3650	1192
9 02_sex_miss	0	1

Note how data display changes after you tick "Tab" in "Separator options". It must be as displayed on the right. Thus after ticking "Tab", accept and the data are now written into the sheet:

	A	B	C
1	rowvartxt	case	noncase
2	01_age_Q1	1014	627
3	01_age_Q2	1365	465
4	01_age_Q3	1367	395
5	01_age_Q4	1312	432
6	01_age_missing	0	2
7	02_sex_fem	1408	728
8	02_sex_male	3650	1192
9	02_sex_miss	0	1

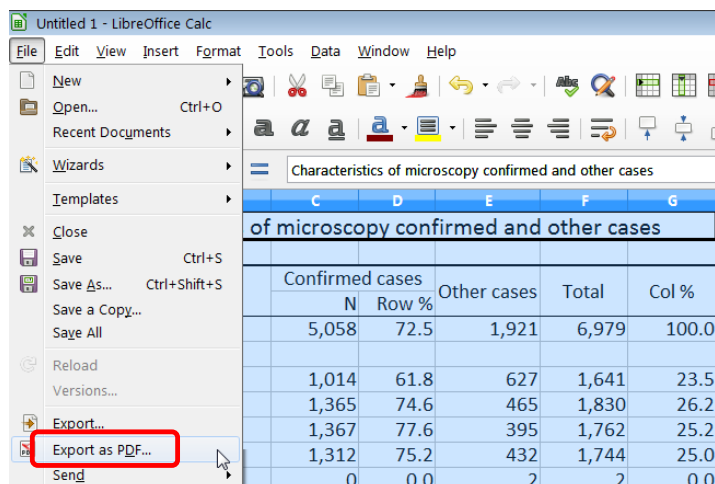
To get just this, we wouldn't have needed the spreadsheet: it is not different from what is already available in the EpiData output. If you look at the left bottom of the spreadsheet you see:



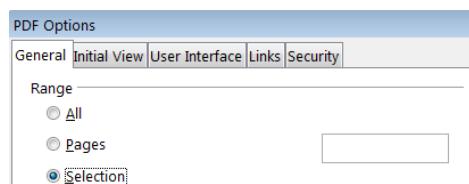
There are two sheets and the currently open is the "input" sheet. If you click on the "output" sheet you get to the formatted edition of the "input" sheet:

	A	B	C	D	E	F	G
1	Characteristics of microscopy confirmed and other cases						
2							
3		Confirmed cases					
4	Characteristic	N	Row %	Other cases	Total	Col %	
5	Total	5,058	72.5	1,921	6,979	100.0	
6	Age quartiles						
7	Quartile 1	1,014	61.8	627	1,641	23.5	
8	Quartile 2	1,365	74.6	465	1,830	26.2	
9	Quartile 3	1,367	77.6	395	1,762	25.2	
10	Quartile 4	1,312	75.2	432	1,744	25.0	
11	Age missing	0	0.0	2	2	0.0	
12	Sex						
13	Female	1,408	65.9	728	2,136	30.6	
14	Male	3,650	75.4	1,192	4,842	69.4	
15	Sex missing	0	0.0	1	1	0.0	
16	Disease site						
17	Pulmonary	5,030	85.6	848	5,878	84.2	
18	Extrapulmonary	23	2.1	1,072	1,095	15.7	
19	Site missing	5	83.3	1	6	0.1	
20	Patient category						
21	New	4,366	74.5	1,491	5,857	83.9	
22	Relapse	502	97.9	11	513	7.4	
23	After failure	64	100.0	0	64	0.9	
24	After default	22	91.7	2	24	0.3	
25	Transfer in	76	52.4	69	145	2.1	
26	Other	8	14.3	48	56	0.8	
27	Category missing	20	6.3	300	320	4.6	

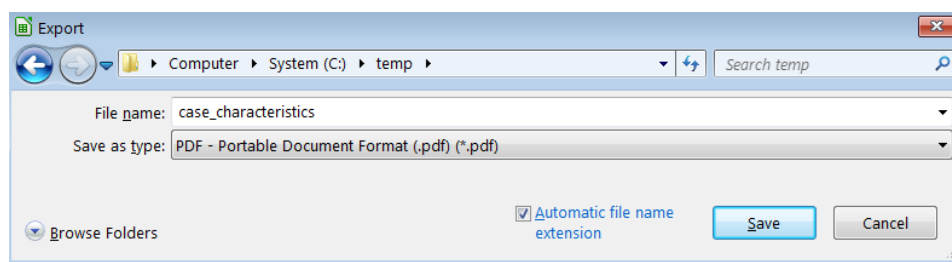
OpenOffice has an inbuilt PDF maker and we will make use of it. Mark the entire data range (cells A1 to G27), then go to “File” and click “Export as PDF”:



Choose “Selection” in the “Range” of the opening window:



then give your file a name and a location where you can easily find it:



When you view it, the properly formatted PDF file looks definitely more appealing than the original output:

Report from the National Tuberculosis Program					
Characteristics of microscopy confirmed and other cases					
Characteristic	Confirmed cases		Other cases	Total	Col %
	N	Row %			
Total	5,058	72.5	1,921	6,979	100.0
Age quartiles					
Quartile 1	1,014	61.8	627	1,641	23.5
Quartile 2	1,365	74.6	465	1,830	26.2
Quartile 3	1,367	77.6	395	1,762	25.2
Quartile 4	1,312	75.2	432	1,744	25.0
Age missing	0	0.0	2	2	0.0
Sex					
Female	1,408	65.9	728	2,136	30.6
Male	3,650	75.4	1,192	4,842	69.4
Sex missing	0	0.0	1	1	0.0
Disease site					
Pulmonary	5,030	85.6	848	5,878	84.2
Extrapulmonary	23	2.1	1,072	1,095	15.7
Site missing	5	83.3	1	6	0.1
Patient category					
New	4,366	74.5	1,491	5,857	83.9
Relapse	502	97.9	11	513	7.4
After failure	64	100.0	0	64	0.9
After default	22	91.7	2	24	0.3
Transfer in	76	52.4	69	145	2.1
Other	8	14.3	48	56	0.8
Category missing	20	6.3	300	320	4.6