EpiTour - an introduction to
EpiData Entry

Dataentry and datadocumentation
Http://www.epidata.dk

© Jens M. Lauritsen, Michael Bruus
& EpiData Association

Version 25th. August 2005

EpiData

EpiData is a windows 95/98/NT based program for:

- Defining data structures
- Simple dataentry
- Entering data and applying validating principles
- Editing / correcting data already entered
- Asserting that the data are consistent across variables
- Printing or listing data for documentation of error-checking and error-tracking
- Comparing data entered twice
- Exporting data for further use in statistical software programs

EpiData works on
Windows 95/98/NT/Professional/2000/XP and Machintosh with RealPc emulator.
Linux based on WINE.

Suggested citation of EpiData Entry program:

Lauritsen JM & Bruus M. EpiData (version). A comprehensive tool for validated entry and documentation of data. The EpiData Association, Odense, Denmark, 2003-2005.

Suggested citation of EpiTour introduction:

Lauritsen JM, Bruus M. EpiTour - An introduction to validated dataentry and documentation of data by use of EpiData. The EpiData Association, Odense Denmark, 2005. Http://www.epidata.dk/downloads/epitour.pdf (See Version above)

This updated version is based on
Lauritsen JM, Bruus M, Myatt M. EpiTour - An introduction to validated dataentry and documentation of data by use of EpiData. The EpiData Association, Odense Denmark, 2001.
For further information and download of latest version: See http://www.epidata.dk

Modfication of this document: See general statement on www.EpiData.dk. Modified or translated versions must be released at no cost from a web page and a copy sent to info@epidata.dk. Frontpage cannot be changed except for addition of revisor or translator name and institution.

Version 25th. August 2005

Introduction and Background

What is EpiData ?

EpiData is a program for DataEntry and documentation of data.

Use EpiData when you have collected data on paper and you want to do statistical analyses or tabulation of data. your data could be collected by questionnaires or any other kind of paperbased information. EpiData Entry is **not** made for analysis, but from autumn 2005 a separate EpiData Analysis is available. Extended analysis can be done with other software such as Stata, R etc.

With EpiData you can apply principles of "**controlled dataentry**". Controlled means that EpiData will only allow the user to enter data which meets certain criteria, e.g. specified legal values with attached text labels(1 = No 2= Yes), rangecheck (only ages 20-100 allowed), legal values (e.g. 1,2,3 and 9) or legal dates (e.g. 29febr1999 is not accepted).

EpiData is suitable for simple datasets like one questionnaire as well as datasets with many or branching dataforms. **EpiData** is freeware and available from Http://www.epidata.dk. A version and history list is available on the same www page.

The principle of EpiData is rooted in the simplicity of the dos program Epi Info, which has many users around the world. The idea is that you write simple text lines and the program converts this to a dataentry form. Once the dataentry form is ready it is easy to define which data can be entered in the different data fields.

If you want to try EpiData during the coming pages make sure you have downloaded the program and installed it.

It is an essential principle of EpiData not to interfere with the setup of your computer. EpiData consists of one program file and a few help files. No other files are installed. (In technical terms this means that EpiData does not install or include any DLL files or system files - options are saved in registry.)

Registration

All users are encouraged to registrate by using the form on www.epidata.dk . By registration you will receive information on updates and help us in decing how to proceed development - and to persuade others to add funding for the development.

Useful internet pages on Biostatistics, Epidemiology, Public Health, Epi Info etc.:

Data types and analysis:	http://www.sjsu.edu/faculty/gerstman/EpiInfo
Statistical routines:	http://www.oac.ucla.edu/training/stata/
Epidemiology Sources:	http://www.epibiostat.ucsf.edu/epidem/epidem.html
Epidemiology lectures:	http://www.pitt.edu/~super1/

Freeware for dataentry, calculations and diagrams:

EpiData (current program) for dataentry is available at www.epidata.dk

Epicalc 2000 Epidemiological oriented calculator. http://www.myatt.demon.co.uk/

EpiGram for drawing flowcharts and diagrams http://www.myatt.demon.co.uk/

OpenEpi Initiative: http://www.cdc.gov/epo/epi/epiinfo.htm

Epi Info home page: http://www.cdc.gov/epo/epi/epiinfo.htm

Steps in the DataEntry Process - principle

1 Aim and purpose of investigation is settled

- Hypothesis described, Size of investigation, time scale, Power calculation ...
- Funding ensured, Ethical commitee ....etc.

2 Ensuring Technical dataquality at entry of data

Collect data and ensure quality of data from a pure technical point of view. Document the process in files and error lists.

- done by applying legal values, range checks etc
- entering all or parts of data twice to track typing errors.
- finding the errors and correcting them

3 Consistent data and logical assertion.

The researcher cross examines the data. Trying to see if data are to be relied upon:

- Sound from a content point of view (no grandmothers below age of xx, say 35)
- Amount of missing data. Some variables might have to be dropped or part of the analysis should question influence on estimates in relation to missing.
- Decisions on number of respondents (N).

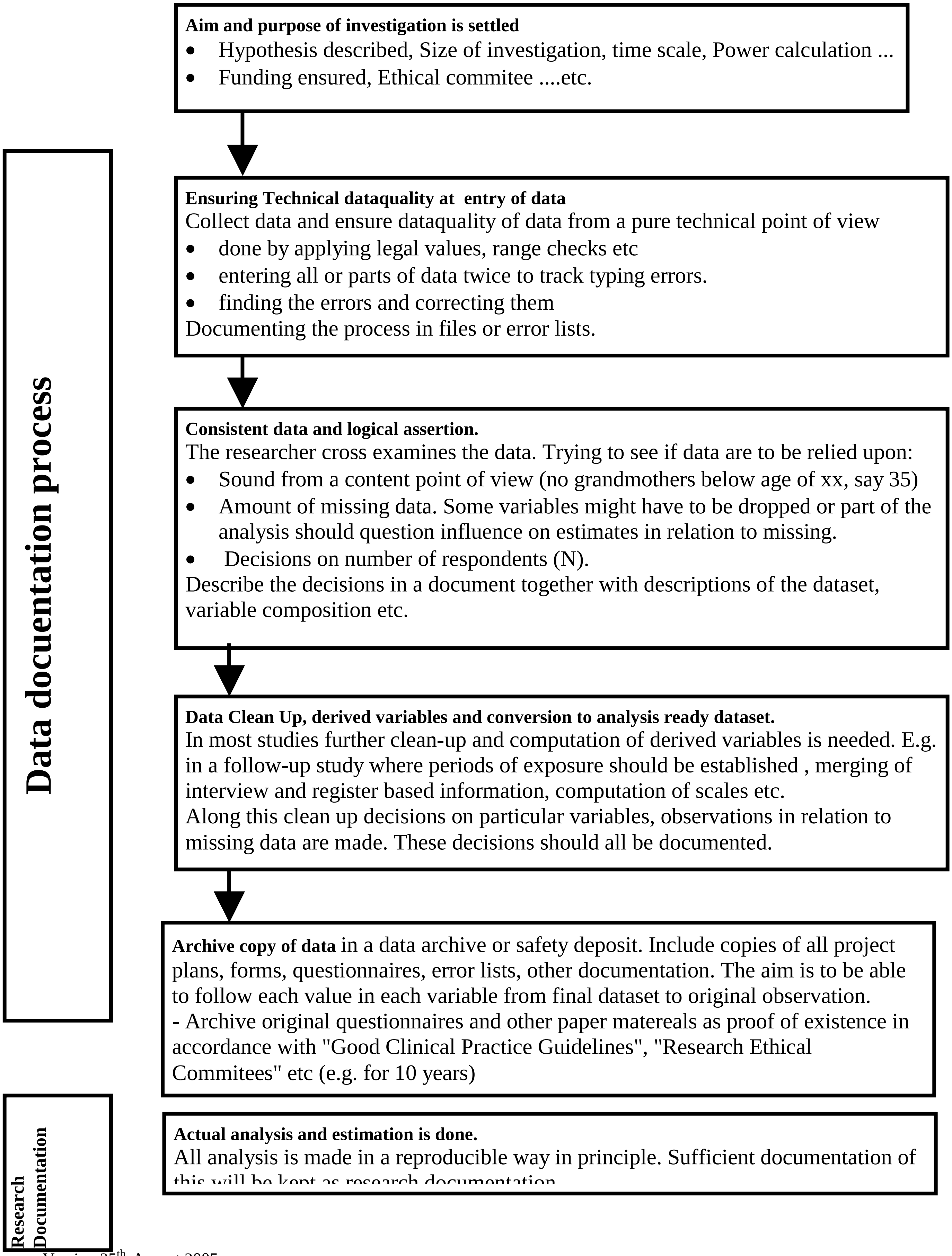
Describe the decisions in a document together with descriptions of the dataset, variable composition etc.

4 Data Clean Up, derived variables and conversion to analysis ready dataset.

In most studies further clean-up and computation of derived variables is needed. E.g. in a follow-up study where periods of exposure should be established , merging of interview and register based information, computation of scales etc. Along this clean up decisions on particular variables, observations in relation to missing data are made. These decisions should all be documented.

5 Archive copy of data in a data archive or safety deposit. Include copies of all project plans, forms, questionnaires, error lists, other documentation. The aim is to be able to follow each value in each variable from final dataset to original observation.Archive original questionnaires and other paper materials as proof of existence in accordance with "Good Clinical Practice Guidelines", "Research Ethical Commitees" etc (e.g. for 10 years)

6 Actual analysis and estimation is done. All analysis is made in a reproducible way in principle. Sufficient documentation of this will be kept as research documentation.



Aim and purpose of investigation is settled

- Hypothesis described, Size of investigation, time scale, Power calculation ...
- Funding ensured, Ethical commitee ....etc.

Ensuring Technical dataquality at entry of data

Collect data and ensure dataquality of data from a pure technical point of view

- done by applying legal values, range checks etc
- entering all or parts of data twice to track typing errors.
- finding the errors and correcting them

Documenting the process in files or error lists.

Consistent data and logical assertion.

The researcher cross examines the data. Trying to see if data are to be relied upon:

- Sound from a content point of view (no grandmothers below age of xx, say 35)
- Amount of missing data. Some variables might have to be dropped or part of the analysis should question influence on estimates in relation to missing.
- Decisions on number of respondents (N).

Describe the decisions in a document together with descriptions of the dataset, variable composition etc.

Data Clean Up, derived variables and conversion to analysis ready dataset.

In most studies further clean-up and computation of derived variables is needed. E.g. in a follow-up study where periods of exposure should be established , merging of interview and register based information, computation of scales etc.

Along this clean up decisions on particular variables, observations in relation to missing data are made. These decisions should all be documented.

Archive copy of data in a data archive or safety deposit. Include copies of all project plans, forms, questionnaires, error lists, other documentation. The aim is to be able to follow each value in each variable from final dataset to original observation.

- Archive original questionnaires and other paper matereals as proof of existence in accordance with "Good Clinical Practice Guidelines", "Research Ethical Commitees" etc (e.g. for 10 years)

Actual analysis and estimation is done.

All analysis is made in a reproducible way in principle. Sufficient documentation of this will be kept as research documentation

DataEntry Process - in practice.

Depending on the particular study the details of the process outlined above will look different. The demands for a documentation based data-entry and clean-up process varies therefore. Let us look at the process in more detail.

a. Which sources for data Based on approved study plans. Decide which sources of data will make up the whole dataset. E.g. a questionnaire, an interview form and some blood samples. Sample/identify your respondents (patients). Generate an anonymous ID variable.

b. Save an ID-KEY file with two variables: **id** and **Social security number**, **Civil registration number** or other appropriate identification of respondents

c. Collect your Data:

questionnaire (common id variable): Enter data with control on variable level of:

- legal values, range, filter questions (jumps), etc.

interview form (common id variable): Enter data with control on variable level of:

- legal values, range, filter questions (jumps), etc.

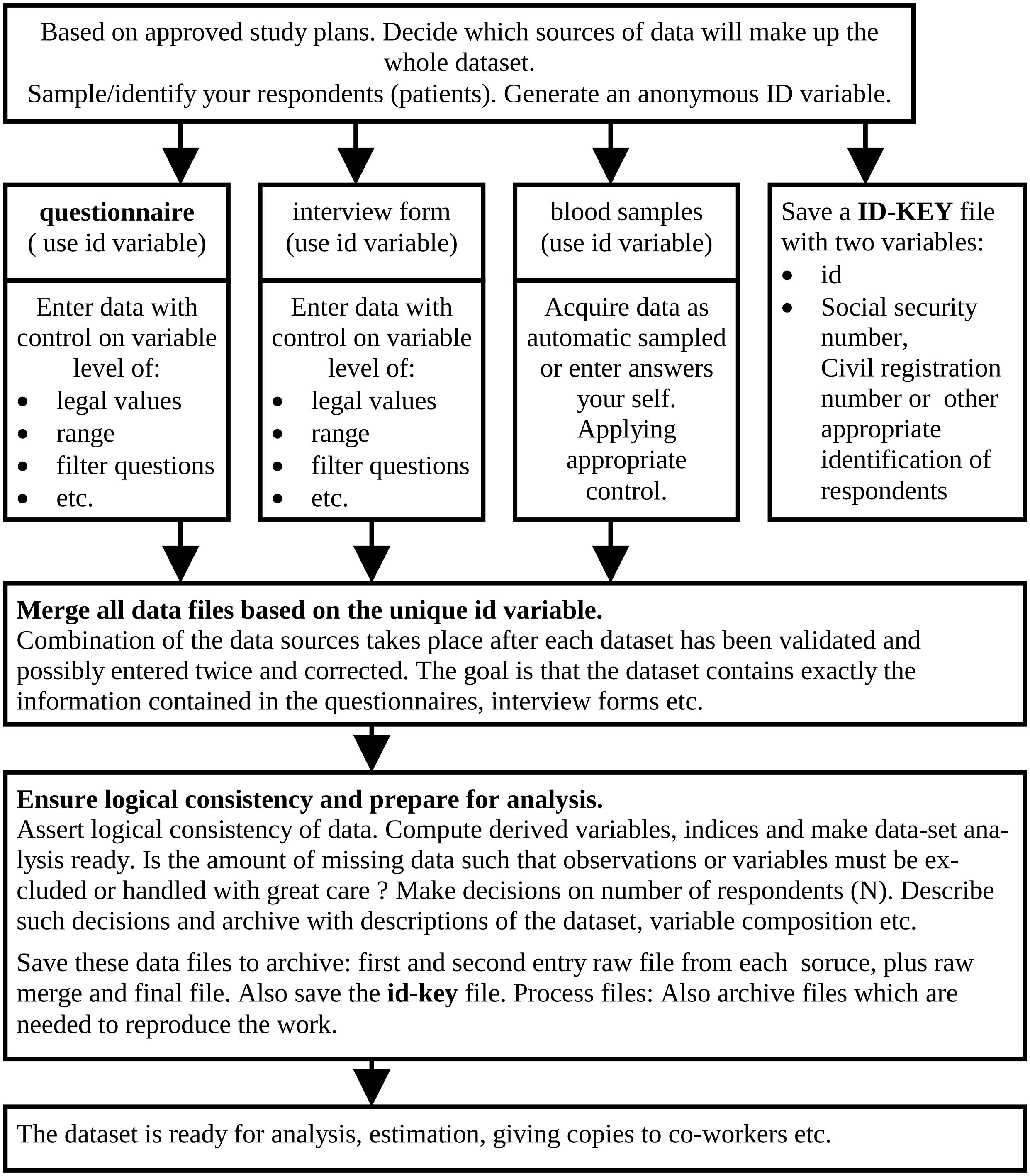
blood samples (common id variable):

- Acquire data as automatic sampled or enter answers your self, applying appropriate control.

d. Merge all data files based on the unique id variable.
Combination of the data sources takes place after each dataset has been validated and possibly entered twice and corrected. The goal is that the dataset contains an exact replica of the information contained in the questionnaires, interview forms etc.

e. Ensure logical consistency and prepare for analysis.
Assert logical consistency of data. Compute derived variables, indices and make data-set analysis ready. Is the amount of missing data such that observations or variables must be excluded or handled with great care ? Make decisions on number of respondents (N). Describe such decisions and archive with descriptions of the dataset, variable composition etc.

Save these data files to archive: first and second entry raw file from each soruce, plus raw merge and final file. Also save the **id-key** file. Process files: Also archive files which are needed to reproduce the work..



Based on approved study plans. Decide which sources of data will make up the whole dataset.
Sample/identify your respondents (patients). Generate an anonymous ID variable.

questionnaire
(use id variable)

Enter data with control on variable level of:

- legal values
- range
- filter questions
- etc.

interview form
(use id variable)

Enter data with control on variable level of:

- legal values
- range
- filter questions
- etc.

blood samples
(use id variable)

Acquire data as automatic sampled or enter answers your self. Applying appropriate control.

Save a **ID-KEY** file with two variables:

- id
- Social security number,
  Civil registration number or other appropriate identification of respondents

Merge all data files based on the unique id variable.
Combination of the data sources takes place after each dataset has been validated and possibly entered twice and corrected. The goal is that the dataset contains exactly the information contained in the questionnaires, interview forms etc.

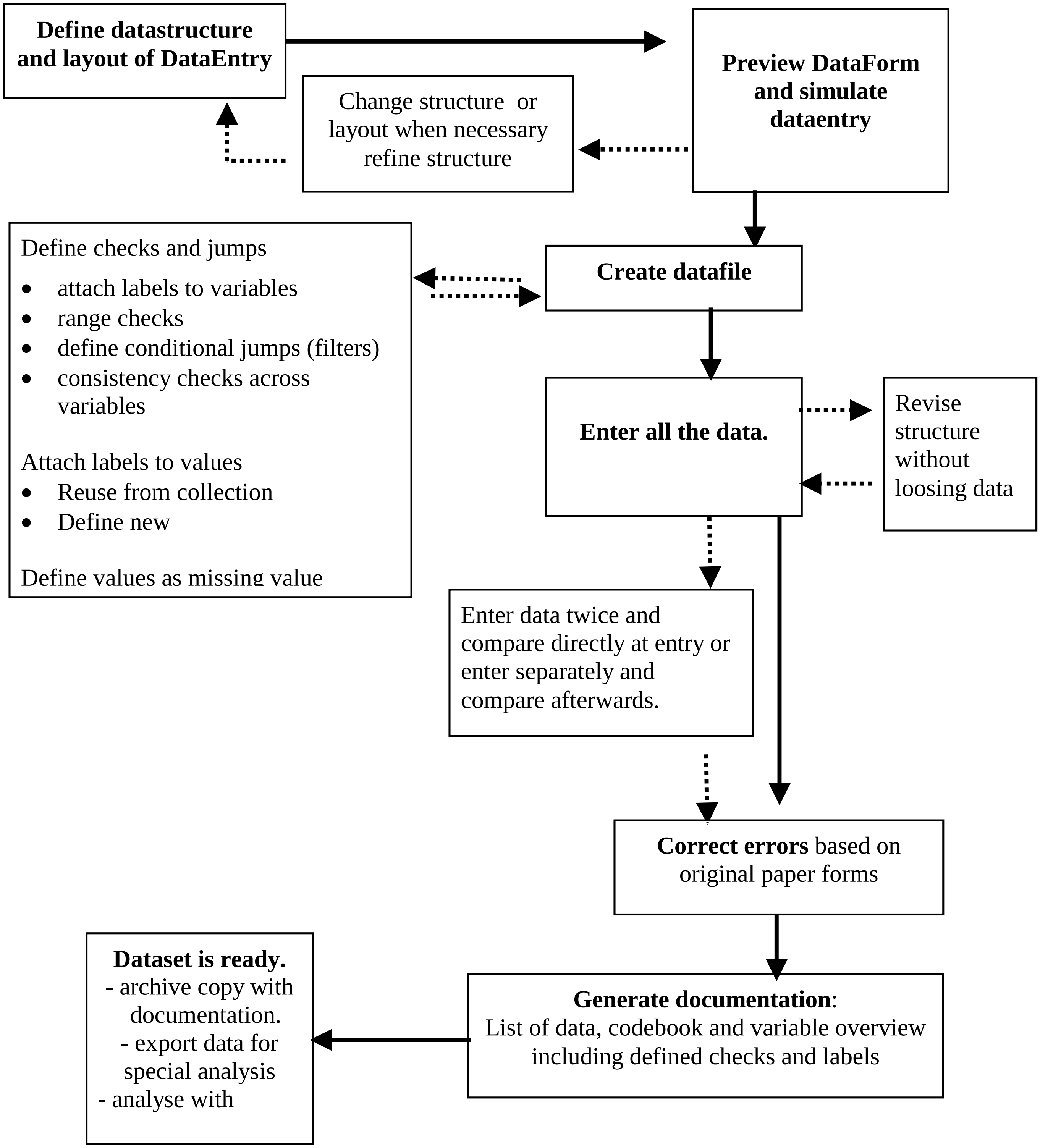
Ensure logical consistency and prepare for analysis.
Assert logical consistency of data. Compute derived variables, indices and make data-set analysis ready. Is the amount of missing data such that observations or variables must be excluded or handled with great care ? Make decisions on number of respondents (N). Describe such decisions and archive with descriptions of the dataset, variable composition etc.

Save these data files to archive: first and second entry raw file from each soruce, plus raw merge and final file. Also save the **id-key** file. Process files: Also archive files which are needed to reproduce the work.

The dataset is ready for analysis, estimation, giving copies to co-workers etc.

Flowsheet of how you work with EpiData Entry

The work process is as this (optional parts are dotted):



Define datastructure and layout of DataEntry

Preview DataForm and simulate dataentry

Change structure or layout when necessary refine structure

Define checks and jumps

- attach labels to variables
- range checks
- define conditional jumps (filters)
- consistency checks across variables

Attach labels to values

- Reuse from collection
- Define new

Define values as missing value

Create datafile

Enter all the data.

Revise structure without loosing data

Enter data twice and compare directly at entry or enter separately and compare afterwards.

Correct errors based on original paper forms

Generate documentation:
List of data, codebook and variable overview including defined checks and labels

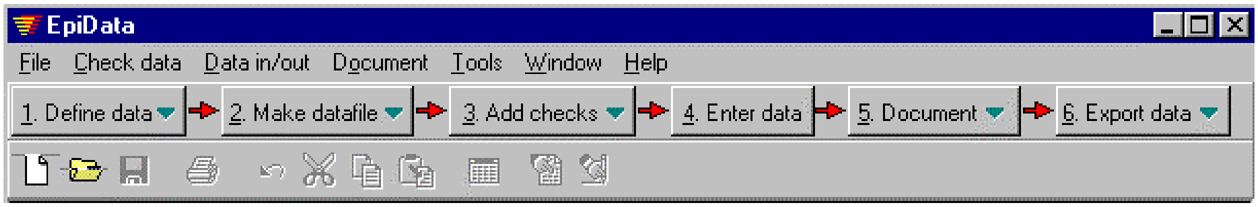
Dataset is ready.
- archive copy with documentation.
- export data for special analysis
- analyse with

Install EpiData

Get the latest version from Http://www.epidata.dk and install in the language of your preference. The installation and retrieval is fast[1] since the whole size of the programme is small (1.5Mb in total).

How to work with EpiData

The EpiData screen has a "standard" windows layout with one menu line and two toolbars (which you can switch off). Depending on the current task, the menu bar changes.



The "Work Process toolbar" guides you from "1. Define data" to "6. Export data" for analysis. The second toolbar helps in opening files, printing and certain other tasks to be explained later.

A. If you want you can switch off the toolbars in the Window Menu, but this EpiTour will follow the toolbar and guide you.

B. Start EpiData now.

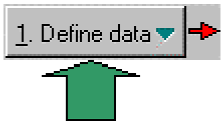
C. Continue by doing things in EpiData and reading instructions in this EpiTour.

D. In the menu Help you can see how to register as a user of EpiData. Registered Users will receive information on updates.

E. Proceed to 1. Define and test DataEntry Form

[1] If you are on a slow modem line you might not agree to "fast", but in comparison to many programmes this is a small size

1. Define and test Data Entry



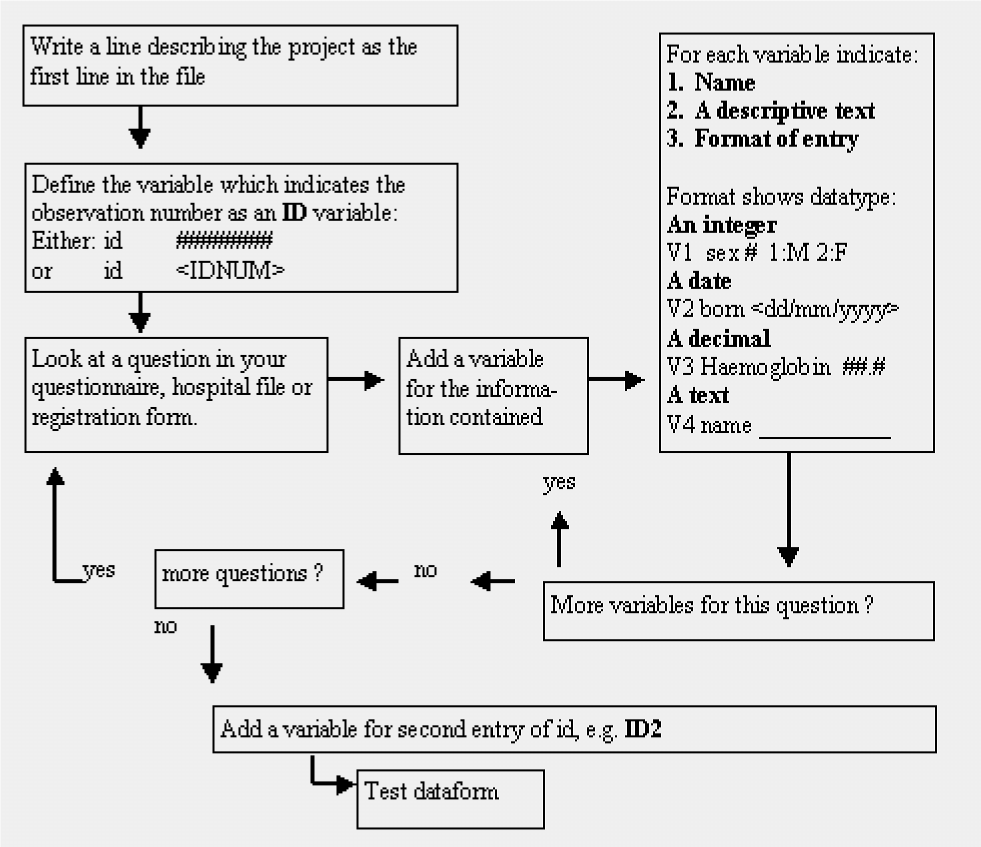
1.. Point at “Define data” part and “new qes file” . An empty file called “untitled” is shown in the “Epi-Editor”. A qes file defines variables in your study. “**Qes**” is an abbreviation of “qustionnaire”, all types of information can be entered with EpiData. Questionnaire is just a common name for all of them.

2.. Save the empty file and give it the name **first.qes**.
you save files on the “file menu” or by pressing “Ctrl+S”. Notice that in the Epi-Editor “untitled” changes to “first.qes”.

Write now in the Epi-Editor the lines shown:
Explanation: Each line has three elements:
A.. Name of variable (e.g. v1 or exposure).
B.. Text describing the variable.
(e.g. sex or "day of birth")
C.. An input definition, e.g. ## for two digit numerical.

```
My first DataEntry Form
id         <idnum >
V1 sex            #
V2 Height (meter)      #.##
v3 Date of birth <dd/mm/yyyy>
s1 Country of Residence
____________________________
s2 City (Current adress)         <a
>
```

Path Diagram - Building a datadefinition ("qes" file)

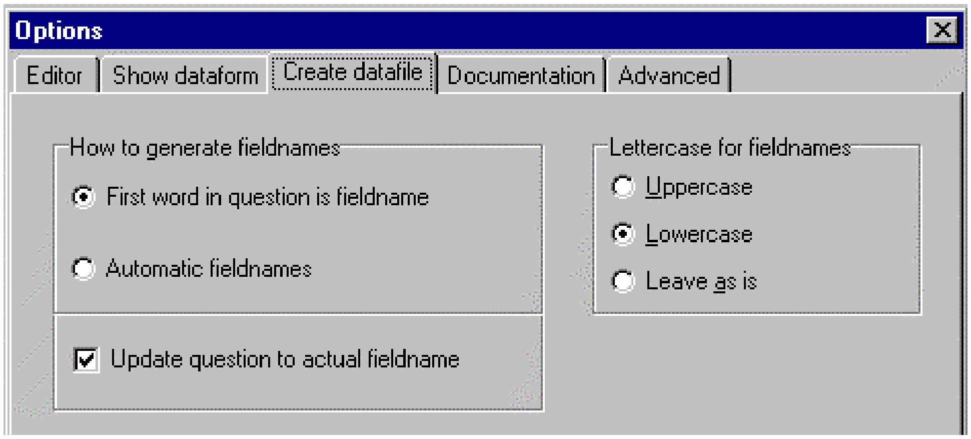

Write a line describing the project as the first line in the file
Define the variable which indicates the observation number as an ID variable:
Either: id ##########
or id <IDNUM>
Look at a question in your questionnaire, hospital file or registration form.
Add a variable for the information contained
For each variable indicate:
1. Name
2. A descriptive text
3. Format of entry
Format shows datatype:
An integer
V1 sex # 1:M 2:F
A date
V2 born <dd/mm/yyyy>
A decimal
V3 Haemoglobin ##.#
A text
V4 name ___________
yes
More variables for this question ?
no
more questions ?
yes
no
Add a variable for second entry of id, e.g. ID2
Test dataform

If you add the descriptive text before the field defining character (e.g. #) then the text will be part of the variable label. If you place it after it will not.

Depending on settings in options, you can get variable names v1, v2 ....v8 or v1age v2sex ... v8Dur in this example:

On next page the options setting is shown. Options are available as part of the "files menu".

```
id <idnum>
V1     Age    ##
V2     Sex    #
V3     Temp   ##.##
V3a    Temp   ##.##
V4     WBC    ##
V5     AB     #
V6     Cult   #
V7     Serv   #
V8     Dur    ##
```



How to generate fieldnames

● First word in question is fieldname

○ Automatic fieldnames

☑ Update question to actual fieldname

Lettercase for fieldnames

○ Uppercase

● Lowercase

○ Leave as is

If you select "first word" as shown in the options (file menu) you get v1, v2.....v8 in the example above.

The field types available in EpiData:

Type	Example	Type	Example
Text	_______	Soundex	<S >
ID-number	<IDNUM>	Boolean	<Y>
Numeric	### ###.##	Tabulator	@<a> @##
Upper-case	<A> <A >	Encrypted	<e>
Date	<dd/mm/yyyy> <mm/dd/yyyy> <yyyy/mm/dd>		
Today's date	<today-dmy> <today-mdy> <today-ymd>		

The tabulator field type is only used to align the other fields. Useful if you use proportionally spaced fonts.

3.. Save the file again as done in point 2. Simplest by pressing Ctrl+s

4.. Now preview the dataform.

Press the toolbar box 2 from left and choose "Preview Dataform" or press Ctrl+T.

On the screen is shown a preview or test DataForm for your **first.qes**. Try moving around with arrows. Notice how the variable types are shown on the status bar. On the preview dataform you cannot save any data, it is designed for testing the number of variables .

If you want other variables switch to the Epi-Editor and add them now.

Press this on the second toolbar for help in formatting fields by pasting them onto the form or Ctrl+Q. you can hide it again by pressing “Esc”.

A different tool can be activated by Ctrl+W or the icon shown. This will "autocomplete" certain parts. E.g. if you add one #,_ , <d, etc and see what happens. Try " and indicate 1.1 or 8.

In date fields **EpiData** will help in completing the date: If you enter ”040599” in a date-dmy field, EpiData will format it as "04/05/1999". If 2-digit years are entered the century will be 1900 for years between 50 and 99 and 2000 for years between 00 and 49. If you enter ”0405”, current year will be used, i.e. in year 2001 saved as ”04/05/2001”. "040503" is saved as "04/03/2003" After entry, all dates are checked. E.g. 29021999 will not be accepted.

Use the tabulator mark @ in front of fields to left align the input field.

When you are satisfied with the DataForm, **close** the form as well as the Epi-Editor (See menu File).

Proceed to next section. Create Datafile

2. Create DataFile

Press **2** (Alt+2) and accept the ”**first.qes**” and ”**first.rec**” names for "make datafile". This will create a physical file on the disk. Preview Dataform only exist in the memory of your computer.

Press **OK** and you have created a datafile based on the definition made earlier:
you have now defined:

A A dataform definition file saved as a file with the name **first.qes**
B An actual datafile which will contain the data, saved as **first.rec**.
Rec is an abbreviation for ”record” or observation.

(If you look at the files in windows explorer you might only see two files with the name first, but not the "qes" and "rec" extension. If this is so setup your explorer to show extensions for all files, you do that in preferences)

Proceed to next section. Add Checks

Color selection and options

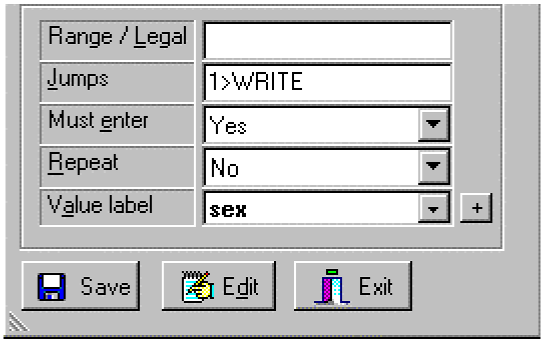
On the menu file one point is "options". If you activate this a number of options can be set. Among others the way the color of the screen dataentry fields are shown. Some options are easy to understand others are for more experienced users. You cannot destroy data by changing options.

Variable names

Variable names can take two forms: e.g. v1sex (8 first characters in sentence-default) or v1 (first word of sentence). If you prefer to use names like v1 v2 v3 .... t1, then you should pick the "Use first word as fieldname" option in "create datafile" (see menu file - "options"). Close all files before changing options.

3 Add checks of DataEntry

When you ”add checks” **you specify rules** for dataentry. **Adding checks is optional**. Move directly to Enter Data if you are not concerned with checks at this point.



Range / Legal	
Jumps	1>WRITE
Must enter	Yes
Repeat	No
Value label	sex

Add checks has five basic aspects and one "advanced":

1.. **label:** Add descriptive text to numerical values (Value label)

2.. Restrict dataentry to certain values **(range, legal)** (and **label)**

3..Specify sequence of dataentry E.g. fill out certain questions for males only, **(jumps)**

4..**Mustenter** you must enter a value

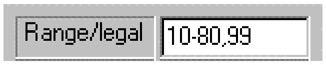
5.. **Repeat** – copy value from previous record. E.g. if you are typing in data from several schools. The class will be the same for several children.

Advanced: Help messages and other extended definitions of computations, if .. then ...endif structures, definition of missing values etc.. This is added with the "Edit" button next to the Exit button. This aspect is NOT covered in this EpiTour, see help file.

A rule of thumb:
Make simple rules
Use range/legal when you have continous data (e.g. from 1 to 100 or 4.2 to 13.1)
Use labels when you have categorial data or a few values (e.g. 1,2 or 1,2,3,4,5,9)

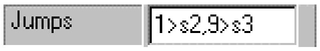
Examples:

range 10-80 plus single value 99



Range/legal	10-80,99

Jumps: On value 1 goto s2: 1>s2 . On value 9 goto s3: 3>s3



Jumps	1>s2,9>s3

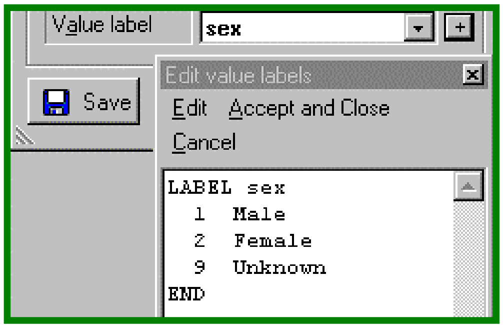
Now add value labels to a variable.

Make sure you have NOT specified any range/legal to that field.

The idea of the label is that it associates an input number with a text description for a numerical field. Or for a short string field it gives a longer description. It makes statistics output more readable and makes sure, that the value entered is interpreted correctly.

EpiData exports information on value labels if you export data to Stata format, and to SPSS and SAS by creation of command files and raw datafiles. Commercially available programs (e.g. StatTransfer) can convert these files further to other statistics programmes

Each label has a name. If a label name is shown in the small window to the right of "Value label" and you press the "+" further to the right, then the values are shown. E.g. sex as seen in the figure.



```
LABEL sex
   1   Male
   2   Female
   9   Unknown
END
```

You can define new labels, e.g. one with the name "size" with contents 1small 2 normal 3 large. To define a new label press the "+" next to the field value label. A small window pops up. After definition of a label it can be attached to as many variables as you like.

You have now added checks to the dataentry in **first.rec** based on the structure you defined in **first.qes.** The definitions of these checks are saved in a file called **first.chk**.

Continue with Enter Data

Special and further notes.

a.

you can press Ctrl+C and save the current definitions in a buffer. If you then move to a different field on the Dataform and press Ctrl+V the definitions copied will be applied in a single keystroke.

b.

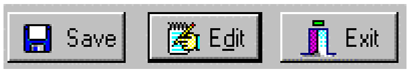
Fieldnames can be added to the jumps automatically. First write the value in the jumps section, (e.g. 1) then add the sign (>) and then click with your mouse on the desired field (Change screen position with PageUp/Page Down keys). The fieldname will be added automatically to the jumps definition.

c

The Edit button gives access to all currently defined checks for the field. Here more complicated structures can be added (e.g. if.. then ...endif) structures, see help file.

Help messages can be defined and shown during dataentry, e.g.:

help "text to show to person inputting data"



d.

For Epi Info users: Most aspects of the Epi Info version 6 check language has been implemented in EpiData.

Proceed to next section. Enter Data

4. Enter Data

Simply activate the **Enter data** on the toolbar and accept **first.rec** for dataentry.

Enter 5-10 records. A record is one observation, one unit in the data.

As seen on the **menu** (Goto) you can jump between records, move to first and last variable, search for values etc. These options can be used if you need to revise data.

The Shift+Del buttons mark a single record as deleted. If something was entered incorrectly and/or messed up, you can mark it for deletion and leave it out while rewriting the file for export. In coming versions an option to permanently delete these records will be implemented.

Sometimes it is useful to enter comments while entering data. While first.rec is open for dataentry, press F8. This will open a small document editor and add the date and time. You can then add some comments. The small file gets the name **first.not** and should be backed up along with the **first.rec**, **first.chk** and **first.qes** files. EpiData can do this for you, see later.

4a. Double Entry of Data

It is good practice to double enter all or part of data and compare directly.

EpiData provides two strategies:
1 Enter data separately in two different files and compare them afterwards.
2 Enter while EpiData compares your entry with a previous entry of the same data.

With both strategies you will have two separate data files, which can be used for documentation.

To initiate double entry look in the **tools menu** and decide which principle you wish to apply.

For strategy 1: Pick the menu : **Copy Structure**

This will allow you to get a copy of the rec file structure plus the check file. Give any new name you prefer for the second dataset.

For strategy 2: Pick the menu : **Prepare Double Entry verification**

EpiData Entry will now create a separate data file xxx_dbl.rec where xxx is the name of your file.

Now enter your data the second time by opening the file you created, either the xxx_dbl from strategy two or the one you saved with strategy one.

Proceed to next section. Document Data

5 Document Data

In this part you can write out information about the structure, date, labels and contents of files and variables. Here you can see an example for the file first.rec, created during the previous parts of this short **EpiTour**.

```
DATAFILE:  C:\data\first.rec
Filelabel: My first datafile is an example

Filesize:          612 bytes
Last revision:     28. okt 2000 22:14
Number of fields:  7
Number of records: 0
Checks applied:    Yes (Last revision 28. okt 2000 22:32)


Fields in datafile:

No. Name Variable label   Fieldtype       Width  Checks            Value labels
-------------------------------------------------------------------------------------
  1 id                    ID-number       6

  2 v1   sex              Integer         1                        sex
                                                                   1: Male
                                                                   2: Female
                                                                   9: Unknown

  3 v2   Height (meter)   Fixed number    4:2    Legal: 0.0-2.30,9

  4 v3   Date of birth    Date (dmy)      10

  5 s1   Country          Text            28

  6 s2   City (Current    Upper-case text 12                       string
         adress)                                                   blt: Baltimore
                                                                   cph: Copenhagen Denm
                                                                   rey: Reykjavik Icela
                                                                   sid: Sidney
                                                                   ndh: New Delhi
                                                                   mom: Mombassa
                                                                   bue: Buenos Aires

  7 t1   Todays Date      Today date-dmy  10

-------------------------------------------------------------------------------------
```

Lists of values can be shown like this:

```
Observation 1

      id                1         v1           Male         v2           1.92
      v3      12/12/1945          s1        denmark         s2      Copenhagen
      t1      28/10/2000
```

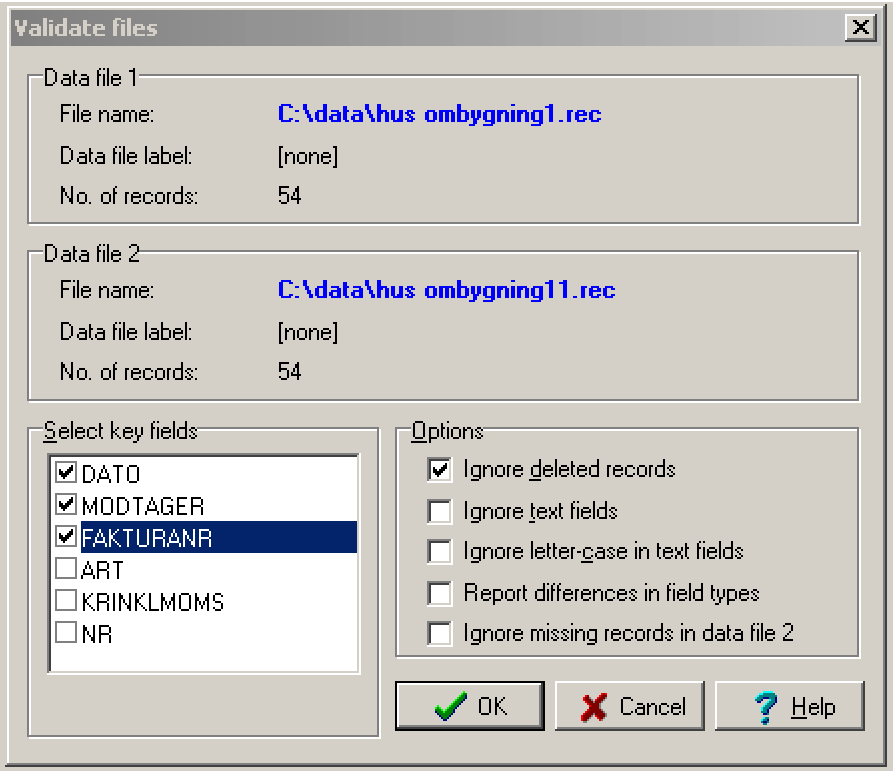
With codebook you can write out condensed frequency tables:

```
Example of this:

v1 ------------------------------------------------------------------------- Age
                   type:  Integer
            range/legal:  0-100

                missing:  0/25
                  range:  [4,82]
          unique values:  22

v2 ------------------------------------------------------------------------- Sex
                   type:  Integer
           value labels:  sex
            range/legal:  1-2,2

                missing:  0/25
                  range:  [1,2]
          unique values:  2

             tabulation:   Freq.    Pct.   Value  Label
                              11    44.0       1  Male
                              14    56.0       2  Female

v3 ------------------------------------------------------------------------ Temp
                   type:  Floating point
            range/legal:  36.00-40.00

                missing:  0/25
                  range:  [36.00,37.50]
          unique values:  12

                   mean:  36,84
               std. dev:  0,37
```

and with "Dataentry notes" you can further edit the file activated by F8 at dataentry.

Validate Double Entry

Choose the validate dublicate files part of the menu. WHen you specify the two files EpiData Entry should know which records two compare. If you have an index or group variable use these. Records can be matched on up to three variables.



Data file 1
File name: C:\data\hus ombygning1.rec
Data file label: [none]
No. of records: 54

Data file 2
File name: C:\data\hus ombygning11.rec
Data file label: [none]
No. of records: 54

Select key fields
☑ DATO
☑ MODTAGER
☑ FAKTURANR
□ ART
□ KRINKLMOMS
□ NR

Options
☑ Ignore deleted records
□ Ignore text fields
□ Ignore letter-case in text fields
□ Report differences in field types
□ Ignore missing records in data file 2

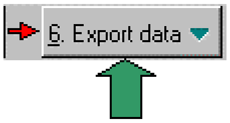
A report will be shown with number of discrepancies.

```
.... extract of full report.....

------------------------------------------------------------------------------
RESULTS OF VALIDATION:
------------------------------------------------------------------------------
Records missing in data file 1:           0
Records missing in data file 2:           0

Number of common records found:          54
Number of fields checked per record:      6
Total number of fields checked:         324

0 out of 54 records had errors ( 0.00 pct.)
0 out of 324 fields had errors ( 0.00 pct.)
```

Clarify any discrepancy by editing the data and comparing with original datasheets and redo the validation.

Proceed to next section. Export Data

6 Export, Analysis and options.



The simplest export is a backup of your data. In case of fire, waterpipe failure, theft or breakdown of computers you MUST have a copy in a different location.

Therefore each day's work ends with making a ”**backup copy**”. Activate ”Export data” and try it out.

If you wish to analyse the data with the EpiData Analysis programme you are ready to do so. The files written by EpiData Entry can be analysed as is.

If you are using a different program for analysis you can export data to one of the types:

Simple export (data and variable names)
1..simple (comma separated) ascii file
2..dbaseIII
3..Excel

Complete export (data and variable names, labels)
4..Stata, SPSS and SAS

Import (data and for Stata also variable names, labels)
1..simple (comma separated) ascii file
2..dbaseIII
3..Stata

For use in other statistical programs commercial software for conversion is available, such as Stat/Transfer (http://www.circlesys.com/) or DBMS/Copy.

This was your first dataset with EpiData. Proceed to Support and About EpiData

Support

If you find errors or bugs when using the program or have suggestions for improval please discuss with the EpiData-list available at http://lists.umanitoba.ca/mailman/listinfo/epidata-list

Sources for support:
1..Read the help file to epidata.
2..Read this epitour document
3..Download from http://www.epidata.dk the epidata help file
and the epitour help file in the format of "pdf", which is easy to print.
4. Basic aspects of epidata follows the Epi Info version 6 manuals. This is
available from the Epi Info site: http://www.cdc.gov/Epi Info/

Unfortunately we do not have resources for support of dataentry questions in general refer these to the EpiData-list available at http://lists.umanitoba.ca/mailman/listinfo/epidata-list

About EpiData

EpiData is a Windows 95/98/NT/2000/XP based program (32 bit) for DataEntry. EpiData works on Mac computers based on emulation and on LINUX versions based on WINE.

EpiData development initiated by Jens M. Lauritsen, MD. PhD, Denmark. Program released as freeware by: The EpiData Association, Odense Denmark. Program design: Jens M.Lauritsen & Michael Bruus.

Suggested citation:

Lauritsen JM & Bruus M. EpiData (version). A comprehensive tool for validated entry and documentation of data. The EpiData Association, Odense, Denmark, 2003-2005.

First version of EpiData released as Lauritsen JM, Bruus M., Myatt MA, EpiData, version 1.0-1.5. A tool for validated entry and documentation of data. County of Funen Denmark and Brixton Health UK. 2001.

Funding and acknowledgements.

An updated list of attained funding is available at Http://www.epidata.dk/funding.htm and further credits and acknowledgements at: Http://www.epidata.dk/credit.htm . International translations made to several languages, see Http://www.epidata.dk
For donations to further development see help file or send an e-mail to info@epidata.dk

Disclaimer

The EpiData software program was developed and tested to ensure fail-safe entering and documentation of data. We made every possible effort in producing a fail-safe program, but cannot in any circumstance be held responsible for errors, loss of data, work time or other losses incurred by or in relation to the program.