

# Data Management II

Sasivimol Rattanasiri, Ph.D  
Clinical Epidemiology Unit  
Ramathibodi Hospital

1

---

---

---

---

---

---

---

---

## Data management processes

Design and create case report form (CRF)

Collect data by using CRF

Design and create database



Specify data quality control



Enter data into database

2

---

---

---

---

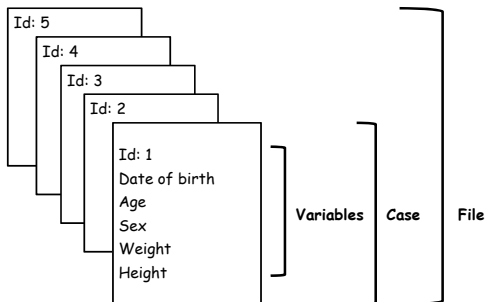
---

---

---

---

## Database file



3

---

---

---

---

---

---

---

---

### Data set for database file

ID	Date of birth	Age	Sex	Weight	Height
1	12/12/1973	37	M	56	167
2	10/11/1988	22	M	78	178
3	03/08/1963	47	F	45	158
4	14/09/1986	24	M	67	169
5	23/10/1981	29	F	41	155

4

---

---

---

---

---

---

---

---

### Database Management System (DBMS)

The DBMS is a set of computer programs which perform a wide range of operations:

- creating new files
- entering new records
- sorting, searching, and editing
- and so on.

5

---

---

---

---

---

---

---

---

### DBMS software package

There are many different DBMS software packages:

- dBase
- Paradox
- EpiData
- Access
- and so on

6

---

---

---

---

---

---

---

---

## Introduction to EpiData

- EpiData is a program for data entry and documentation of data only.
- However, you can export the data to a number of data formats for statistical analysis.

7

---

---

---

---

---

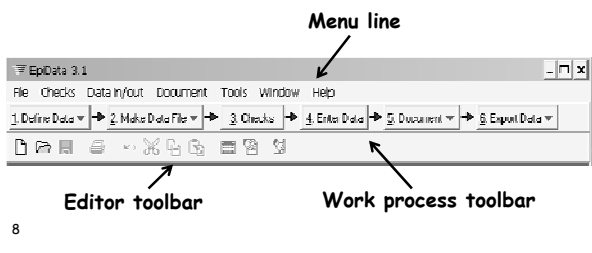
---

---

---

## Overview of EpiData

The EpiData screen has a standard windows layout with one menu line and two toolbars.



8

---

---

---

---

---

---

---

---

## Process of creating database file with EpiData

Define data -> Questionnaire file (.QES)

Make data file -> Record file (.REC)

Add/revise checks -> Check file (.CHK)



9

---

---

---

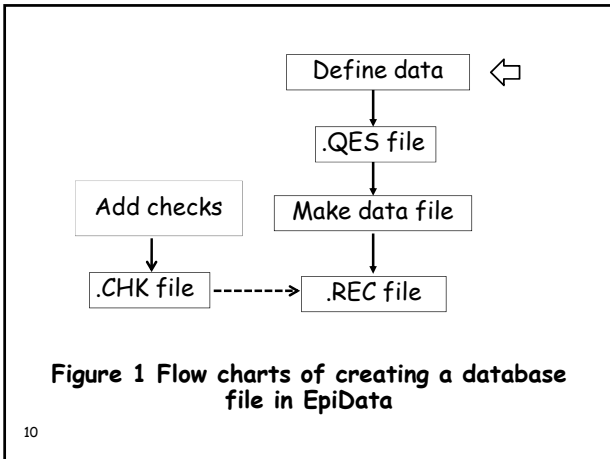
---

---

---

---

---




---

---

---

---

---

---

---

---

**1. Define data**

The first step is to define the structure of a database file by writing types of information for each variable

- Variable name ↩
- Variable label
- Variable type
- Variable length

11

---

---

---

---

---

---

---

---

**Variable names**

- Must not exceed eight characters.
- Must not contain any space or punctuation marks.
- Has to begin with a letter, not a number.
- Can contain any sequence of letters and digits.
- Can be upper or lower case.

12

---

---

---

---

---

---

---

---

### Variable names

- Names should describe the variable they refer to, e.g. sbp is more suitable name for a systolic blood pressure variable than var5.
- For large CRF, it may be easier to use question numbers, e.g. Q5, Q6, Q7, as variable names.

13

---

---

---

---

---

---

---

---

### Examples of illegal variable names

#### Variable name

1date	Begins with a number
Last name	Contains a space
countryoforigin	Longer than 8 letters

14

---

---

---

---

---

---

---

---

### 1. Define data

The first step is to define the structure of a database file by writing types of information for each variable

- a) Variable name
- b) Variable label ⇐
- c) Variable type
- d) Variable length

15

---

---

---

---

---

---

---

---

### Variable labels

- The variable label is the meaning of the variable name.
- When you are working with lots of variables, you should explain the meaning of all variable names in labels.
- For example, variable label of dateb is date of birth.

16

---

---

---

---

---

---

---

---

### 1. Define data

The first step is to define the structure of a database file by writing types of information for each variable

- a) Variable name
- b) Variable label
- c) Variable type ←
- d) Variable length

17

---

---

---

---

---

---

---

---

### Variable types

The variable type indicates the characteristic of the variable such as

- Text,
- Numeric,
- Logical,
- Date,
- Auto ID, etc.

18

---

---

---

---

---

---

---

---

### **Text variables**

- Text variables are used for holding data consisting of letters and/or numbers.
- You can enter numbers into text variables but you cannot perform mathematical operations with them.

19

---

---

---

---

---

---

---

---

### **Numeric variables**

- Numeric variables are used for holding numerical information.
- Numeric variables can be used for holding categorical and continuous data.
- Numeric variables can be defined to hold either integers or real numbers.

20

---

---

---

---

---

---

---

---

### **Date variables**

- Date variables are used for holding dates.
- You can perform simple arithmetic such as addition or subtraction one date variable from another date variable.

21

---

---

---

---

---

---

---

---

## Examples of variable types

Name	Type
ID	Numeric
Date of birth	Date
Age at enrollment	Numeric
Sex	Numeric
Do you have any underlying diseases?	Numeric
Specify medications	Text

22

---

---

---

---

---

---

---

---

## 1. Define data

The first step is to define the structure of a database file by writing types of information for each variable

- a) Variable name
- b) Variable label
- c) Variable type
- d) Variable length ⇐

23

---

---

---

---

---

---

---

---

## Variable length

- The length of a variable defines how much data it can hold.
- A text variable with length 10 will be able to hold up to ten letters or numbers.

24

---

---

---

---

---

---

---

---



## Variable length

- A numerical variable with length 3 will be able to hold numbers between -99 and 999.
- The length of a variable must correspond to the maximum anticipated number of letters and/or numbers.

25

---

---

---

---

---

---

---

---

## Specify variable type and length

Type	EpiData definition
Text	-----
ID-number	<idnum>
Numeric	### or ###.##
Upper-case	<A> or <A >
Logical	<Y>
Date	<dd/mm/yyyy> <mm/dd/yyyy>
Today's date	<today-dmy>, <today-mdy>

26

---

---

---

---

---

---

---

---

Data Registry for Ramathibodi Renal Transplant Center  
 Part I Recipient form

```

id          ID ###
sexr       Sex of recipient #
dateb      Date of birth <dd/mm/yyyy>
dater      Date of transplantation <dd/mm/yyyy>
age        Age at transplantation ##
ht         Height (cm) ###.#
wt         Weight (kg) ###.#
bmi        Body mass index ##.##
  
```

↑ Variable name                      ↑ Variable label                      ↑ Variable type

recipient.qes  
Pos 1 Line 1  
27

---

---

---

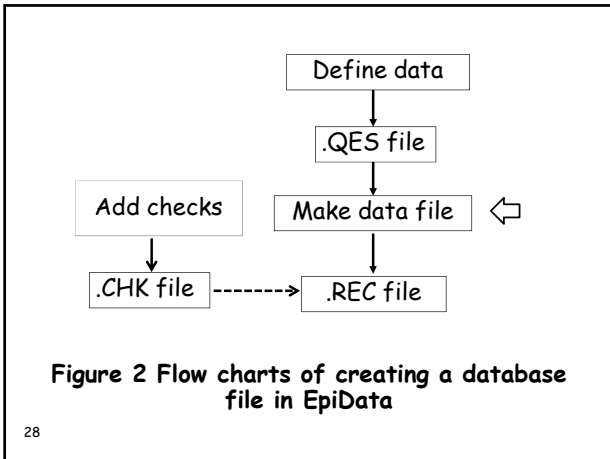
---

---

---

---

---




---

---

---

---

---

---

---

---

## 2. Make data file

- The second step is to create the database file based upon the database structure.
- The make data file function is used to create a record (.REC) file from questionnaire (.QES) file.

29

---

---

---

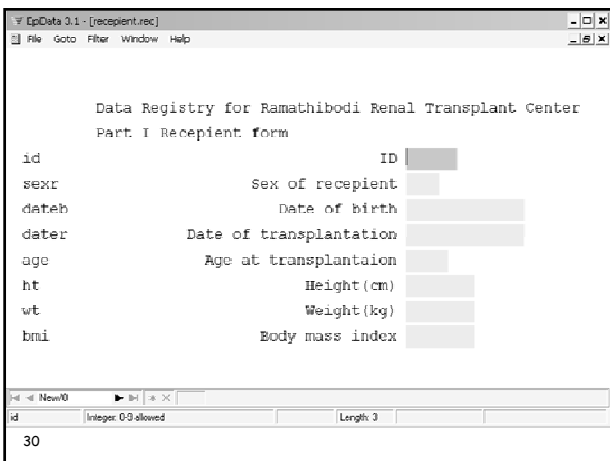
---

---

---

---

---




---

---

---

---

---

---

---

---

## Summary

At the end of this step, you can enter the data set into the database file.

31

---

---

---

---

---

---

---

---

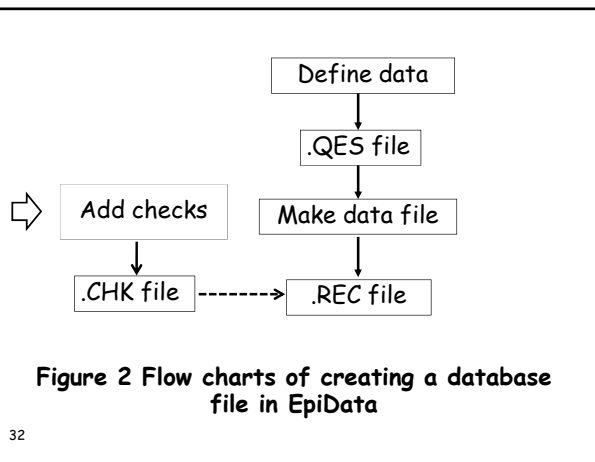


Figure 2 Flow charts of creating a database file in EpiData

32

---

---

---

---

---

---

---

---

## Interactive checking functions

EpiData provides functions that allow you to do data interactive checking such as:

- Must enter variables
  - Range and legal values
  - Attach value labels to variables
  - Repeated variables
  - Conditional jumps
  - Programmed checks
- Basic checks
- Advanced checks

33

---

---

---

---

---

---

---

---

### **Must-enter variables**

- You can specify that a certain variable must be filled with a value other than leave it blank.
- The variable can be left blank unless it is defined as a must-enter variable.

34

---

---

---

---

---

---

---

---

### **Legal values**

- You can specify the legal values for a certain numeric variable.
- The input must match one of a specified list of values.

35

---

---

---

---

---

---

---

---

### **Range checks**

- You can specify that a certain variable must lie between two limited values.
- You may mix range checks and legal values checks (e.g. for missing value codes).

36

---

---

---

---

---

---

---

---

### **Attach value labels to variables**

- You can specify a valid value label set for categorical variables such as sex, blood group.
- When you specify a value label set, you specify both the legal values for a variable and the meaning of each of the legal values.

37

---

---

---

---

---

---

---

---

### **Repeat variables**

- You can specify that a certain variable on a new record will automatically keep the value from the previous case.
- This is useful for data that seldom changes.

38

---

---

---

---

---

---

---

---

### **Conditional jumps**

- You can specify a value for a variable that will cause the entry to jump to a target variable.
- If the tests fail the entry moves to the next variable.

39

---

---

---

---

---

---

---

---

### Interactive checking functions

EpiData provides functions that allow you to do data interactive checking as:

- Must enter variables
  - Range and legal values
  - Attach value labels to variables
  - Repeated variables
  - Conditional jumps
  - Programmed checks
- 
- The diagram shows a list of six items. A bracket on the right side of the first four items is labeled 'Basic checks'. An arrow points from the fifth and sixth items to an oval labeled 'Advanced checks'.

40

---

---

---

---

---

---

---

---

### CONFIRM

- After a field has been filled, the cursor automatically moves to the next field.
- If CONFIRM is set, the cursor will move to the next field after the enter key is pressed.
- The command CONFIRM is placed in the before file block.

41

---

---

---

---

---

---

---

---

### TYPE COMMENT

- This command can be used in fields which specify value labels for the variables.
- If TYPE COMMENT is set, after a value is entered in a field, the text connected to the value is written to the right of the field.
- The command TYPE COMMENT is placed in the field block.

42

---

---

---

---

---

---

---

---

### **TYPE COMMENT**

- The syntax can be defined as:  
`TYPE COMMENT [color]`
- An option color code can be added to specify a different color than the default blue.
- Valid colors are Red, Purple, Green, Yellow, Black, White, etc.

43

---

---

---

---

---

---

---

---

### **TYPE COMMENT**

- A `TYPE COMMENT ALLFIELDS` is special form of `TYPE COMMENT` which is applied to all fields which specify value labels for variables.
- The syntax can be defined as:  
`TYPE COMMENT ALLFIELDS [color]`
- This command is placed in the before file block.

44

---

---

---

---

---

---

---

---