Part A. EpiData Entry

Part A: EpiData Entry

Exercise 1: A simple Questionnaire

At the end of this exercise you should be able to:

a. Define the different types of fields/variables (text, numeric, date) and know when to use them.

b. Create a data documentation sheet from a simple questionnaire

Like Epi Info 6, EpiData Entry uses the same principle of what we call the `QES-REC-CHK` (pronounced "Ques-Rec-Check") files principle.

First we create a questionnaire (a form defining the fields) from which we then create a data entry file, and finally we create a so-called Check file linked to the data entry file to control data entry.

But let us proceed step by step. Let us say we have the following questionnaire:

Laboratory serial number: ___

Date specimen received (dd/mm/yyyy): ___/___/_____

Sex: ___

Age in years: ____

Reason for examination: ___

Result of specimen 1: ___

Result of specimen 2: ___

Result of specimen 3: ___

This might present a typical simple questionnaire as used by an interviewer. Often such questionnaires are first completed on paper. This is actually an excerpt from the Tuberculosis Laboratory Register proposed by The Union:

Tuberculosis Programme Form 2

Tuberculosis laboratory register Year_____

Lab Serial No.	Date specimen received	Name	Sex M/F	Age	Name of referring facility	Address - patient for diagnosis	Reason for examination*		Results of specimen			Only for SS+ for diagnosis: TB Number or treatment centre**	Remarks
							Diagnosis (tick)	Month of follow up	1	2	3		

We will use this register as the basis for this course. For the time being, you plan to write a short and concise computer questionnaire, retaining only variables that are easy to capture and are likely to be useful for the analysis.

Each of the questions can be conceived of as a variable and the answer to the question as the value that the variable takes for a particular individual. We will give each variable a unique field name. A completely entered questionnaire for one individual is called a record. We will later enter 15 records (one each for each individual), each with eight fields (corresponding to the number of questions in the questionnaire).

For the time being, we will use field names that consist of one single word that has ***not more than ten characters***.

> Note that some other analysis software may accept only a field length of eight characters. If you later plan to export your EpiData files for analysis to such a software package and you had used the full field length of ten, then your field names get truncated.

You may verify the set-up in "File" "Options" "Create Data File". The field name we use might be chosen in a way that it has some meaning relating to the question. There are different types of entry fields for the variables (we will follow the EpiData Entry notation and call them "Fields"):

Text fields: These fields take letters or numbers or a combination of these as possible values, like `PETER`, `KOCH1882`, `giraffe`, `45677` etc. If you enter a number into such a field you will not be able to make any calculation with it. These fields are also sometimes designated as character or alphanumeric fields.

Numeric fields: These are numbers. The numbers might be integers like `885`, `33`, `1235` or real numbers like `3.4`, `6.88`, `66.5` (also called floating). You can make calculations with such fields.

Date fields: In different countries, different ways of writing dates are used and this can be confusing for people from another culture. Some write *5 March 2005*, others *March 5 2005*, and again others *2005 March 5*. EpiData Entry lets you choose the type of date you wish to take. In this course we will use European dates, i.e. dates of the format *5 March 2005* or symbolized with `DD/MM/YYYY`.

One other type of variables is called "logic" or "Boolean" variables. This is sometimes used in food-borne outbreak investigations. There, answers to questions on food items eaten is limited to "yes" and "no" and "missing". In EpiData Entry, this type of field accepts only the values `Y`, `N`, `0`, and `1`. There is no need for using this additional type of field. It is easy to circumvent by using numeric fields with a label block, and we actually discourage the use of the field type as this is a field type which might pose problems in analysis.

While you are asked to limit the length of the field name, you have much more flexibility with the length of the value a field can take (up to a field length of 80), but we will try to make an as efficient use as possible, that is we will limit the value length to the minimum needed.

Data Documentation Sheet

It is good practice to write what we call a **data documentation sheet** before you make your actual EpiData Entry QES file. EpiData Entry refers to this as **Codebook**.

Note: Field names cannot exceed a length of ten characters, and must be a single word not several words separated by spaces (the space counts). The Field name "`Date of birth`" would be truncated to "`DATE`" (which is a reserved name), thus better use "`DOB`": ensure sticking to single words.

Note: If the Field label begins with a word that is identical to the Field name, you will note later in EpiData Analysis, that this word will be truncated from the Field label. For instance, if your Field name was `SEX`, and you used `SEX OF EXAMINEE` as your Field label, this would be truncated to `OF EXAMINEE`. While this can be fixed easily in EpiData Analysis, it is preferable to prevent it by choosing an alternative Field label during questionnaire design.

This is how we would write such a data documentation sheet:

Field name	Field label	Field type	Field length	Field values	Value labels	Comment
serno	Laboratory serial number *	I	4	1,…,9000 9001, 9002,…		Serial number starting with 1 each year Reserve and assign these numbers sequentially if serial number is not unique, and write a data entry note (use F5 to open a note file)
regdate	Registration date	D	10	01/01/2000,…,31/12/2005 01/01/1800		Range of legal registration dates No or incomplete date provided
sex	Examinee's sex	T	1	F M 9	Female sex Male sex Sex not recorded	

**** Note***: Commonly, it will be preferable to make the identifier a text field. If it is a number, as in this case here with the laboratory serial number, precautions must be taken to distinguish e.g. "0001" from "1", requiring that the numeric value is entered into one field, and another field, the actual identifier field, is automatically correctly calculated to add leading zeros where appropriate.

Task:

- ***Complete the data documentation sheet for all fields in the questionnaire. Note that you should always define a value if no answer was provided to a question.***