

Epi Info Beginner's Manual with Exercises



Great Lakes Epidemiology Center
Community Based Research Training

Updated: 02/2004

Epi Info Beginner's Manual with Exercises

Table of Contents

Introduction to Epi Info	4
Information and Support	4
Overview of Epi Info	4
Why Use Epi Info?	4
How to Run Epi Info	5
Epi Info Components	5
General Terminology	6
The Survey Process	6
Installing Epi Info	7
Exercises, Examples, and Notes	8
Exercise 1: Creating a New Project	8
Create a new questionnaire	8
Create new fields/variables	8
Field names	9
Field and variable types	9
Data entry check code options	10
Move fields	10
Rename and add pages	10
Add group of fields/variables	11
Create legal values	11
Program skip patterns for yes/no fields	12
Exiting <i>MakeView</i> screen	12
Opening/modifying an existing view	13
Customize alignment grid	13
Setting manual tab order	13
Program age calculation	14
Program BMI calculation	14
Program skip patterns for text fields	14
Exercise 2: Entering Data	15
Opening an existing project	15
Adding data to a questionnaire	15
Navigating through questionnaire	16
Finding and editing records	17
Exercise 3: Basic Data Management and Analysis	19
Opening analysis screen	19
Reading or opening a project to analyze	20
Sorting and selecting records	20

Exercise 4: Basic Data Analysis	21
Frequencies/counts	21
Means	21
Tables	21
Routing output	22
Customizing output	23
Exercise 5: Calculating Anthropometric Measures	24
Exercise 6: Reading and Writing Different Formats	25
Reading/importing an Excel table	25
Reading/importing an Epi file	25
Exercise 7: Intermediate Data Management & Analysis	26
Defining new variables	26
Assigning values to new variables	26
Saving program files (.pgm)	27
Opening and running programs	27
Defining and recoding ages to age groups	28
Saving changes into a new data table	28
References	29

I. Introduction to Epi Info

A. Information and Support

CDC Epi Info Hotline (Free Assistance)

Telephone: 770-488-8440
Fax: 770-488-8456
E-mail: epiinfo@cdc.gov
Web site: www.cdc.gov/epiinfo/

GLITC EpiCenter

Telephone: 715-588-3324
Fax: 715-588-3607
E-mail: glitc@glitc.org
Web site: www.glitc.org

B. Overview of Epi Info

Epi Info (hereafter referred to as Epi Info) is a public domain database and statistics program for use by public health officials (doctors, nurses, epidemiologists, etc.) managing databases for public health activities, conducting outbreak investigations, and performing statistical applications. Epi Info allows the user to develop a questionnaire, customize the data entry process, enter data, and analyze the data. Statistics, graphs, tables, and maps can be produced with simple commands. **Epi Info is free, downloadable software provided by the CDC** (www.cdc.gov/epiinfo/). It can also be obtained on CD-ROM. This guide is designed for use by persons who may only have basic computer skills and will only use this software periodically, giving step-by-step instructions for basic tasks. This manual also attempts to correlate Epi Info 2002 commands with those of the Epi 6 DOS version.

C. Why use Epi Info?

Epi Info allows for a database to be created, from which data can be analyzed in an easy manner that spreadsheet programs (e.g. Microsoft Excel) cannot perform. For example, if you want to calculate sums, means, and ranges, Excel will be sufficient. However, if you want to calculate something more complicated, such as the number of males between the ages of 15 and 24 who answered no to question #17, then using Epi Info would be very practical. It is easy to use in a wide range of questionnaire-based inquiries, such as satisfaction surveys, community needs assessments, and program evaluations.

This manual will help new users easily maneuver through the most basic components of Epi Info, such as:

- setting up data entry screens, called "Views"
- inserting codes to assist in checking the data entry process
- entering data and editing records
- performing simple data analyses.

D. How to Run Epi Info

Once Epi Info is installed on your computer, the easiest way to “run” the software is clicking the Epi Info icon on your desktop (the first screen that appears after the computer is turned on). The Epi Info main menu should then appear:



(An interesting note: the image on the main menu is a famous map that was used to identify the source of a cholera outbreak in London, 1846).

At the bottom of the main menu, Epi Info Version 3.2 should appear.

E. Epi Info Components

1. *MakeView*

This screen is the first step to making a questionnaire (also called a form or “View”), which will automatically create a database. Like a word processor, fields or questions can be typed onto the screen to be filled in later on the data entry screen. Data entry instructions to *MakeView* contain the same functions as EPED and CHECK in Epi 6.

2. *Enter Data*

This is the data entry program for Epi Info. Data is entered into the form or “View” created in *MakeView* and constructs a data table that is stored in the database. *Enter Data* is equivalent to ENTER in Epi 6.

3. *Analyze Data*

This program allows access to data in Epi Info “Views” or data tables to perform analyses. It can produce lists, frequencies, cross tabulations, and other epidemiologic statistics, such as odds ratios, relative risks, and p-values (consult a statistics textbook for further study). Graphing and mapping are also available using this component (not covered in this manual). *Analyze Data* is equivalent to ANALYSIS in Epi 6.

4. *Nutrition*

This program compares data on height, weight, age, sex, and arm circumference with international reference standards for assessment of nutritional status. The *Nutrition* component can easily calculate Body Mass Index (BMI) values and BMI percentiles for persons under 19 years old.

5. *EpiMap* (not covered in this document or training)

This is a program for creating Geographic Information Systems (GIS) maps and overlaying survey data onto these maps.

F. General Terminology

Project: An Epi Info file is also referred to as 1) Epi Info database and 2) .MDB files (Microsoft Access). A project consists of a View and Data Table

View: Contains information about the data entry screen's appearance, such as field names and field types.

Data table: Stores the actual data values for the corresponding View.

G. The Survey Process: From Development to Presentation

1. Develop, pretest, and administer a survey.
2. Construct a "View" (data entry program) using the *MakeView* function.
3. *Enter Data* into a prepared "View" to compile a data table.
4. *Analyze Data*.
5. Interpret the data, write a report, and present key findings.

II. Installing Epi Info

You will need to uninstall any earlier versions (Epi Info 2000 or Epi Info 2002) that may be on your system.

- a. On your desktop, click **Start**, then **Programs**, then **Epi Info 2002 (or 2000)**, then **Uninstall Epi Info 200X**.
- b. Continue to follow the instructions on your screen.
- c. When this is complete, install the new Epi Info, Version 3.

A. Installation from the Internet

Step 1: Log on to CDC's website at www.cdc.gov/epiinfo/epiinfo.htm

Step 2: Click the **Download** button to perform a Web Install of the latest Epi Info, Version 3.2

Step 3: Save the file to a temporary folder on your hard drive (anywhere except where Epi Info will be stored).

Step 4: Go to this temporary folder and double click the **Setupweb.exe** icon for a complete installation by following the directions on your screen. You **must be** connected to the Internet while this installation occurs. Web Install will NOT result in a copy of setup files after the installation is complete. This means that you will not be able to install the program onto another system with these files. To do so, you must download the Complete Installation Package, then save onto a CD-ROM.

Step 5: Begin using Epi Info

B. Installation from a CD-ROM

Step 1: Insert disc in CD-ROM drive.

Step 2: Open **My Computer**, click on **(E:)**

Step 3: Click on **Epi Info** folder and select **Full Version** folder

Step 4: Click on **Setup** icon. The Installation Wizard should begin. Follow the instructions as they appear. Most of your selections should be **OK** and **Next**.

Step 5: Click **OK** to put Epi Info icon on your desktop. Make sure icon is on your desktop; double click to make sure it works.

Step 6: Begin using Epi Info.

III. Exercises, examples, and notes

Exercise 1: Creating a new Epi Info “Project” (or database) by creating a new “View” (or questionnaire/form)

Objectives: At the end of this exercise, participants should be able to:

- Create a new “project”
- Add a new “view” to the “project”
- Add variables to the “view”
- Set properties to variables
- Create legal values and codes for data entry
- Add and rename pages to “view”
- Save the “view”
- Change the “view”/questionnaire design
- Add groups of variables to the “view”

Step 1: Create a new questionnaire (Also known as a form or “View”)

- a. Start Epi Info by clicking on the **Epi Info** icon on the desktop.
- b. Click on the button labeled **Makeview** from the main Epi Info 2002 screen.
- c. Click on **File** and choose **New**.
- d. When the **Create or Open PROJECT Dialog** box is displayed, enter a **File Name** (Example: *Obstetrics*) then click **Open**. This is your new Database/Project (which is in Microsoft Access format indicated by the .mdb suffix). Other examples could be “Youth Obesity Study”, “or “Diabetes Survey”.
- e. Enter a **View Name** (Example: *Prenatal*) as a subset or data table of your Database. Then click **OK**.
- f. **MakeView** screens and subsequent work is automatically saved by the program OR to be safe you can click **File** and select **Save**.

Step 2: Creating fields/variables in the questionnaire

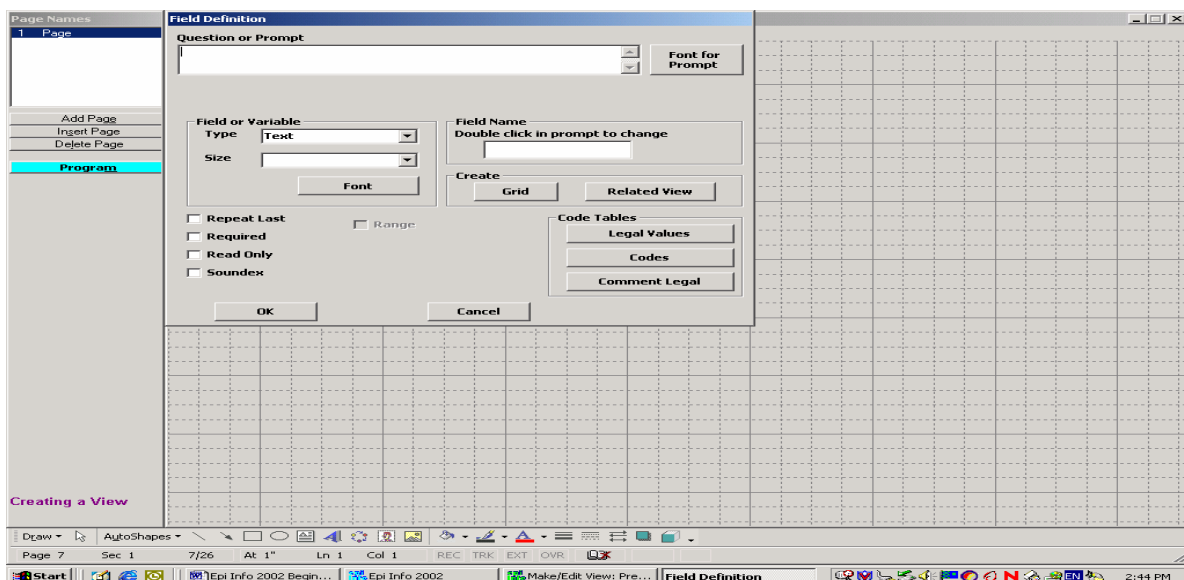
- a. Right click on the **Make/Edit View** screen about where you want the field to be displayed. A **Field Definition** dialog box appears.
- b. Enter text in the **Question or Prompt** box. The text entered will appear on the questionnaire screen. The **TAB** key moves the cursor to the next box.
- c. Click the **Type** scroll down arrow to select the type of variable you want. Some field types will ask for a certain **Pattern** to be selected. Font size and style can be manipulated by clicking on the **Font for Prompt** box. Click **OK**.
- d. For example, refer to the table below, type in the ‘Question or Prompt’, choose a field or variable type and complete the field definitions. Click **OK**. Remember to right click on **Make/Edit View** screen to add each new field/variable. More details about Variable Types, Field Definitions, and Field Names are explained after the table.

Question or Prompt	Field or Variable Type	Field Definitions	Field Name
Obstetrics and Gynecology	Label/Title	Bold Italic, Size 18	Obstetrics&Gyn
Patient Information	Label/Title	Bold, Size 14	PatientInformation
Patient ID	Number	Pattern ##	ID
Last Name	Text		LastName
First Name	Text		FirstName
Health Facility	Text	Select "Repeat Last"	Facility
Address	Multiline		Address
Phone Number	Phone Number	###-###-####	PhNumber

e. **Field Names** are unique "variable names" used when analyzing the data.

- Field names may have a number in them, but cannot start with a number
- Field names may not have symbols or spaces in them
- Field names should be logical and easy to recall for later analyses

Field Names are formed when entering text in the "**Question or Prompt**" box, but can be edited. Before you go to the **EnterData** program, make sure you are happy with the field names, because once the data entry screen is opened, field names cannot be changed unless the data table is deleted (DANGEROUS!).



- ✓ Each data entry item (or variable) is called a "field" in the database.
- ✓ Field or Variable Types summary:
 - **Label/Title** - For display only, they do not hold data.
 - **Number** - Receives numeric data (must use this field type for later calculations, such as averages). Can select number patterns, such as #, ##, ###, ##.##
 - **Text** - Receives text, numbers (cannot use these numbers for later calculations), or symbols.
 - **Multiline** - For more than one line of text, e.g. addresses, comments.
 - **Phone Number** - Special field in phone number format.
 - **Date** - Special field in various date formats.

- **Text (Uppercase)** - Converts text entries into capital letters to ensure correct analysis later. For example, to avoid confusing 'female' with 'Female', this function will convert all to 'FEMALE' and analyze as such.
- **Check Box** - Used with "check all that apply" questions.
- ✓ Data Entry Check Code Options in "**Field Definition**" to help reduce data entry errors and automate the data entry process (not all covered in this manual or training):
 - **Required** - Prevents missing values
 - **Repeat Last** - Automatically repeat the last value entered in that field
 - **Read Only** - Nothing can be typed in this field, used when programming a calculation, such as age
 - **Range** - Sets minimum and maximum values
 - **Legal Values** - Indicates acceptable values, used with Text fields (See *Step 8*)
 - **Comment Legal Values** – Similar to Legal Values, but only a code is saved during data entry and displayed in data analysis. For example, 1 to represent males and 2 to represent females.
- ✓ Check codes are used to check for errors, perform calculations (such as age), and guide the data entry process. Check codes can be written to be very simple or very complicated.
- ✓ Clicking on the blue button labeled **Program** activates the check code editor.
- ✓ Check code must be associated with an existing variable, page, view, or record. The commands will be executed when the data entry cursor enters or leaves the chosen field. The list of variables is available from the drop-down box **Choose field where action will occur**. Variables are organized in pages. Remember that you will not see the question or prompt but only the current variable name.

Step 3: Moving fields

Left click on the question or prompt of the newly created field, don't click on data entry box. While holding down the mouse button, move the cursor to the new location.

Step 4: Rename the current page

- a. Place the cursor on the left side of the screen on the text line **1 Page** and right click.
 - b. Type the title, such as **General Info**, in the box.
 - c. Click **OK** when done.
- ✓ Epi Info can create multiple pages in a single view. Each page can resemble a page in a paper questionnaire or can be used to organize data.

Step 5: Add a new page to the questionnaire

- a. To add a new page, click on the button labeled **Add Page** on the left-hand side of the screen.
- b. Add the desired fields to the view

c. To continue our example, enter the following data:

Question or Prompt	Field or Variable Type	Field Definitions	Field Name
Personal Information	Label/Title	Bold, Size 18	PersonalInformation
Date of Birth	Date	MM-DD-YYYY	DOB
Age	Number	##, Select "Read Only"	Age
Due Date	Date	MM-DD-YYYY	DueDate
Maternal leave?	Yes/No		MatLeave
Marital Status	Text	Select "Required"	MaritalStatus
Smoker?	Yes/No		Smoke
# of Cigarettes (per day)	Number	##	CigsSmoke
Alcohol Consumption?	Yes/No		Alcohol

d. After all the fields are created, name the second page, for example *Personal Info*.

Step 6: Add a group of fields/questions

- ✓ Epi Info can create groups of fields for easy manipulation in analysis and logical organization of the questionnaire.
 - a. Left click above and to the left of the fields wanted in the group.
 - b. Drag the cursor to the lower right corner and release the mouse button. A dotted line box will appear and should outline the desired location for the group box.
 - c. Click on the **Insert** menu and then choose **Group**. Create a name for the group (which will be displayed on the center of the group box). *For example, group the variables **Smoke** and **CigsSmoke**, then name the group: **Tobacco Use**.*
 - d. Optional: choose a color by clicking on **Set Color**.
 - e. Click **OK**.

Step 7: Edit a field

- a. To edit a field, right click on the field name you want to change.
- b. Make changes.
- c. *Example: Change **Patient ID** Field name from **ID** to **PatientID***

Step 8: Create legal values

- ✓ Adding **Legal Values** is the easiest way to customize data entry. It creates a table of allowed entries from which the user can choose.
- ✓ Quality control is maintained for the database while speeding up the process of data entry. **Legal values can be created only on text variables.**
 - a. Right click on field name (*Example: **MaritalStatus***) to get to "**Field Definition**" box.
 - b. Click on the button **Legal Values** located in the lower right side of the pop-up window.

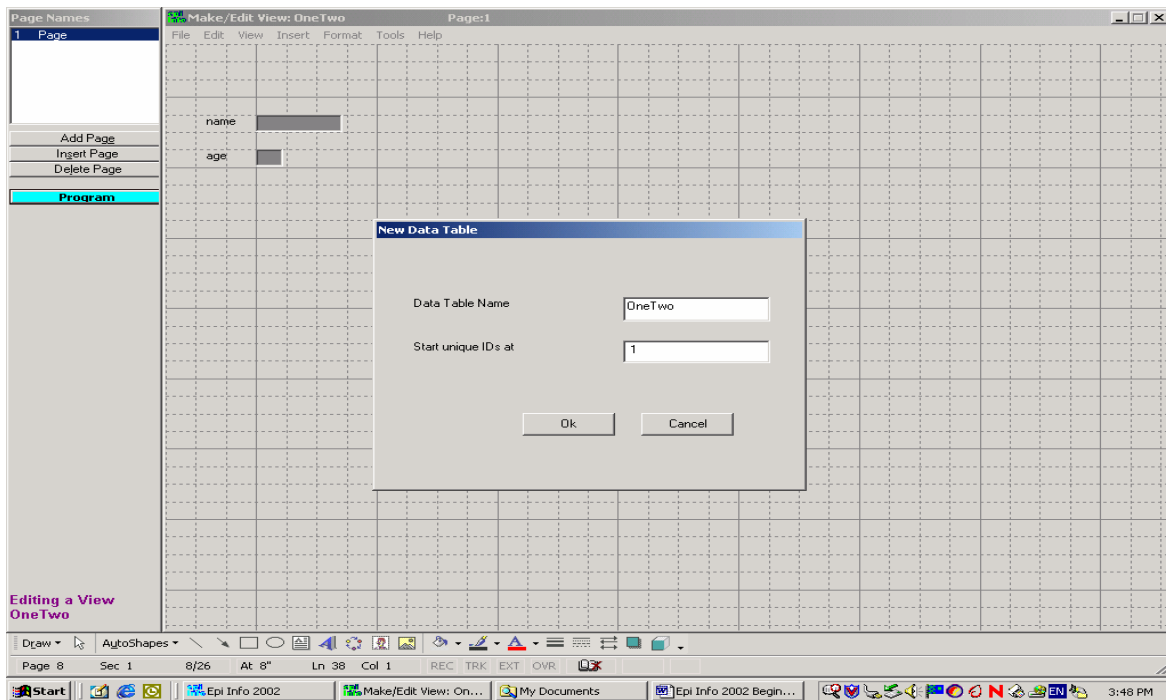
- d. You can use a pre-existing table or you can create a new one. Click on the **CreateNew** button.
- e. Enter value on each line for *MaritalStatus* as: **Single, Married, Divorced, and Other**. If the order of entry is the desired order of display, click on **Do Not Sort**, otherwise the list will be sorted alphabetically.
- f. Click **OK** to save the Legal Values and then click on **OK** again to save the field properties.

Step 9: Programming skip (jump) patterns for the data entry screen (Using a Yes/No Field)

- ✓ Skip patterns refer to skipping over a question based on the answer to another question. For example, if a respondent never smokes, it would make no sense to ask how many cigarettes he smokes each day.
- a. Click the **Program** button
 - b. In the **Choose field where action will occur** box, click the down arrow and select desired field. (For our example, choose **Smoke**)
 - c. Select the **After** option.
 - d. Under the **Records** tab, click **If** button
 - e. Click the down arrow under **Available Variables** and choose **Smoke**, then click **=**, then click **"No"**. The **If Condition** box should now have **Smoke="No"**
 - f. Click the **Then** button.
 - g. Under the **Fields** tab, click the **GoTo** button, then from the down arrow in **Available Variables** choose **Alcohol**.
 - h. Click **OK** until back to questionnaire screen.

Step 10: Exiting MakeView

When exiting the **MakeView** screen, you will see a **New Data Table** box. Do **NOT** click OK unless all the field names have been finalized. If all field names are finalized, click **OK**. If all field names are not finalized, then click **Cancel**.



If you accidentally click OK before you have finalized all Field Names, you can delete that field in the view, and then re-enter it with the desired field name.

Step 11: Modifying a view or Opening an existing project

- Click on the **Makeview** button from the main Epi Info screen.
- Click on **File** and choose **Open**.
- When prompted, choose the project containing the view you want to modify or open (Example: Obstetrics).
- Enter or choose a **View Name** (Example: Prenatal) as a subset or data table of your Database. Then click **OK**.

Optional Step: Customize the Alignment Grid (not covered in this training)

- ✓ To customize the grid, click on **Format** and then on **Settings** from the pull-down menu.
- ✓ To remove the gridlines, switch **Visible grid on screen** check boxes to off. Note the other available settings.
- ✓ By default, **MakeView** will snap all fields to a grid. You can turn off the grid, change the spacing, or remove the visible gridlines from the screen.

Optional Step: Manual tab order (not covered in this training)

- ✓ The order of data entry is controlled by the order of creation or modification of the fields, NOT by the position of the fields on the screen.
- ✓ To manually change the tab order, click on **Edit** and then **Order of FieldEntry (TabOrder)**. To change the tab order, click on the desired field, and choose either the **Up** or the **Down** button. Then click on **OK**.

Optional Step: Program calculations (not covered in this training)

- ✓ Some survey participants are not comfortable giving their age directly, but age can be calculated by subtracting current date/survey date from the date of birth.
 - a. Click the **Program** button
 - b. In the **Choose field where action will occur** box, click the down arrow and select **Dob** field. Make sure the **After** button is clicked so that Age will be calculated after Dob data is entered.
 - c. Click on **Assign** in the command tree on the left side of the screen.
 - d. In the **Assign Variable** box, choose **Age**, since this is the variable receiving the result of the computation.
 - e. In the **= Expression** blank, type **YEARS(Dob, Systemdate)**. Systemdate automatically enters the date stored in the computers system. You could also enter DateofSurvey here or whatever other variable that was used in the questionnaire.
 - f. Click **OK** when done and then click **Save** in the Program Editor window. Then click **OK** at the top of the screen to exit from the program.
- ✓ You could have the program automatically calculate a person's Body Mass Index (BMI) given the person's weight and height (not covered in this training).
 - a. IF **weight** and **height** were variables in the questionnaire, then you could make an assigned variable of **BMI** and use the expression **=weight/(height*height)** to calculate BMI
 - b. Click **OK**

Optional Step: Programming a skip pattern for the data entry screen (Using a Text Field)

- a. Click the **Program** button.
- b. In the **Choose field where action will occur** box, click the down arrow and select desired variable (for example: MaritalStatus). Make sure the **After** button is checked.
- c. Click the **Records** tab, then click the **If** button.
- d. Fill in the **If condition** box with variable and values with quotation marks (since this is a text field). (For example: MaritalStatus="1").
- e. Click the **Then** button.
- f. Click the **Fields** tab, and then click the **GoTo** button.
- g. Select the variable to skip to from list under **Available Variables**.
- h. Click **OK** until back to the questionnaire screen.

Exercise 2: Entering Data

Objectives: At the end of this exercise, participants should be able to:

- Enter data
- Navigate through programs and records
- Use the find/search feature

Step 1: Opening an existing project to enter data

- ✓ There is more than one way to open a project, depending if your program is already open or if you are opening it anew. The following is only one such pathway.
 - Click on **Enter Data** from the Epi Info main menu and select **File** and **Open**. From the dialog box, identify the project created in Exercise 1 (**Obstetrics**), click on it, and then click on **Open**.
 - Select view **Prenatal**, then **OK**.
 - The first time you access the **Enter Data** program for a new view, you will get a message that indicates: **There is no Data Table for this view. A new Data Table will be created**. Click **OK**. Another box titled **New Data Table** will open, click **OK**.
- ✓ The **Enter Data** option can also be accessed in the *MakeView* screen by clicking on **File** then choosing **Enter Data**.
- ✓ Remember: once a new Data Table has been created, you cannot return to the **MakeView** program to modify field names without doing some additional, sometimes risky manipulations.
- ✓ After creating a new View with skip patterns and calculations, some people like to make sure that all of their programs work properly by entering some sample data. This data is then not wanted. You can delete this data table by clicking on **Utilities** on the Main Menu, then **Visualize Data**, then click **File, Open Epi Info Project**, then the name of your project, select **Tables**, then right click on the view (or table) you want to delete. Select **Delete**. You **CANNOT** retrieve this deletion once you have completed it!!

Step 2: Adding data to the questionnaire

- Type the following data for the first patient in the **Prenatal** view (use **TAB** or **ENTER** to advance to the next field/variable)

1 Page - General Info

Patient ID:	01
Last Name:	Smith
First Name:	Jane
Health Facility:	Northside Hospital
Address:	123 Main St
Phone:	404-555-1234

2 Page - Personal Info

Date of Birth:	12/08/1965
Age:	37
Due Date:	01/06/2000
Maternal Leave?:	Yes
Marital Status:	Married
Smoker?:	Yes
# of Cigarettes:	12
Alcohol Consumption:	Yes

- c. The cursor should automatically advance to the next record after entering data into the last field, but you can also advance by clicking the **NEW** button.

1 Page - General Info

PatientID: 02
 Last Name: Jones
 First Name: Mary
 Health Facility: Northside Hospital
 Address: 456 Lake Way
 Phone: 404-555-2345

2 Page - Personal Info

Date of Birth: 05/21/1968
 Age: 35
 Due Date: 11/29/1999
 Maternal Leave?: Yes
 Marital Status: Single
 Smoker?: No
 # of Cigarettes: 0
 Alcohol Consumption: No

1 Page - General Info

PatientID: 03
 Last Name: Steer
 First Name: Anita
 Health Facility: Northside Hospital
 Address: 893 Jake Lane
 Phone: 404-555-3456

2 Page - Personal Info

Date of Birth: 07/24/1975
 Age: 28
 Due Date: 12/07/1999
 Maternal Leave?: Yes
 Marital Status: Married
 Smoker?: Yes
 # of Cigarettes: 34
 Alcohol Consumption: No

1 Page - General Info

PatientID: 04
 Last Name: Lee
 First Name: Jean
 Health Facility: Grady Hospital
 Address: 34 Lake View Way
 Phone: 404-555-9213

2 Page - Personal Info

Date of Birth: 08/02/1980
 Age: 23
 Due Date: 08/19/1999
 Maternal Leave?: Yes
 Marital Status: Single
 Smoker?: Yes
 # of Cigarettes: 5
 Alcohol Consumption: Yes

1 Page - General Info

PatientID: 05
 Last Name: Kent
 First Name: Eloise
 Health Facility: Northside Hospital
 Address: 5741 Bay Circle
 Phone: 404-555-5643

2 Page - Personal Info

Date of Birth: 07/02/1980
 Age: 23
 Due Date: 09/21/1999
 Maternal Leave?: Yes
 Marital Status: Married
 Smoker?: No
 # of Cigarettes: 0
 Alcohol Consumption: No

Step 3: Navigating through the questionnaire

- ✓ On the left-hand side, under the **Record** section, click on the arrows to navigate through the entered records.
- ✓ The << sign brings the data-entry screen to the first record, while the < sign brings the data-entry screen to the previous record.

- ✓ Conversely, the > brings the data-entry screen to the next record, and the >> brings the data-entry screen to the last record.
- ✓ To go directly to a specific record number, click in the white box, type in the record number and click on **Enter**.

The screenshot shows the 'Obstetrics and Gynecology' Patient Information form in Epi Info 2002. The interface includes a menu on the left with 'General Info' and 'Personal Info'. The main form area has input fields for Patient ID, Last Name, First Name, Health Facility, Address, and Phone Number. Navigation buttons at the bottom include '<<', '<', '>', '>>', and 'NewRecord'. The status bar at the bottom indicates 'Revision 2' and 'January 15, 2003'.

Step 4: Finding a record

- a. On the left-hand side, click on the **Find** button. A **Find Record** screen appears with a list of all available fields.
- b. Click on the **LastName** field, and a blank field will appear. Type **Steer** and click **OK**.
- c. A grid appears with the correctly found record. Double-click on any column to bring the cursor to the data-entry screen of that record.
- d. To exit **EnterData**, go to **File** and choose **Exit**. Data entered is automatically saved.

Optional Step: Editing records and making modifications

- a. **Save data** – Once a record has been retrieved and modified, click **Save data** button to save the changes.
- b. **Mark record as deleted** – This button only marks records for deletion, it doesn't actually delete the record from the data table – it just excludes the record from data analyses.

- ✓ Once data exists, proceed with caution when modifying views and data tables.
- ✓ It is ok to add a field to your view and lengthen a field.
- ✓ Be careful with removing fields from your view (all data entered will be deleted for this field).

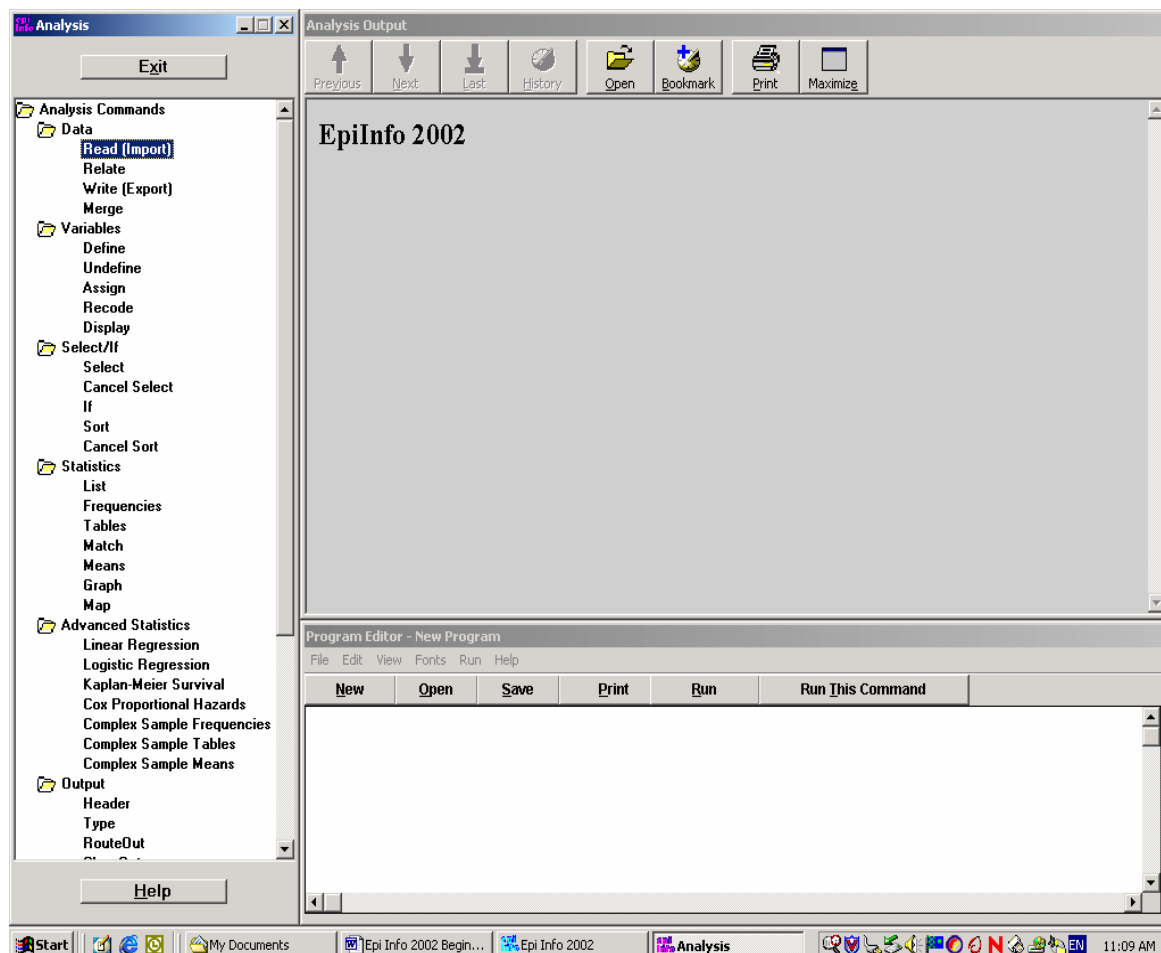
Exercise 3: Basic Data Management and Analysis

Objectives: At the end of this exercise, participants should be able to:

- Open the data analysis function of Epi Info 2002
- Use basic data management functions (Read, Select, Sort, List)

Step 1: Opening Analysis

- To run Analysis, click the **Analyze Data** button on the main menu.
- ✓ Note that all commands are shown in the tree view on the left side of the screen, called Command Generator.
 - ✓ Clicking on a command will bring up a dialog. Responding to the questions and clicking **OK** generates and executes a program command automatically in the program editor at the bottom of the screen.
 - ✓ Results appear in the **Analysis Output** window above the program editor



Step 2: Reading an existing project

- ✓ The first command to be used in Analysis is **Read (Import)** – this will allow you to select a project and data table to analyze.
 - a. Click on **Read (Import)**.
 - b. If the **Current Project** box does not reflect the correct project, then click on the **Change Project** at the bottom of the dialog box (avoid using the ... button). Find the project **Obstetrics** and open it. Then from the list of views, choose **viewPrenatal**, and click on **OK**. This view contains the data table that we are interested in.

Step 3: Obtaining a line listing

- a. Click on the **List** command to create a line listing.
- b. Click **OK** with the **Grid** selection checked. This will display all of the variables in the data table along with each individual record.
- c. The asterisk (*) represents all variables available in the database. To list only selected variables, replace the asterisk by the name of the variable(s) to be listed. Note that you can also display “All Except” the listed variables by selecting this option.
- d. **Allow Updates** allows you to make permanent changes to the data table.

Step 4: Sorting the line listing (not covered in this training)

- ✓ To improve readability of your list, you may want to sort records. The **Sort** command will temporarily order the line listing in numeric or alphabetic order by one or more variables.
 - a. Sort the database by **Age** (double click on **Age**) in ascending order (++), and then use **List** to confirm the results.
 - b. To restore data to its original order, click the **Cancel Sort** command, then **OK**.
- ✓ **Sort** is active until the user cancels it or a new file is read.

Step 5: Selecting a subset of records

- ✓ The **Select** command is used to temporarily analyze a subset of the original file.
 - a. To analyze only women who are married in the database, Click the **Select** command and enter **MaritalStatus="Married."** in **Select Criteria** box. Click **OK**.
 - b. Use **List** to show that only three records remain active:

Current View: C:\EPI_Info\Obstetrics.Mdb:viewPrenatal
 Select: MARITALSTATUS = "Married"
 Record Count: 3 (Deleted records excluded) Date: 08/08/2001 3:48:19 PM

- c. To remove the selected criteria, click **Cancel Select**, and then click **OK**.
- ✓ **Select** is active until the user cancels it or a new file is read.

Exercise 4: Basic Analysis (Frequencies, Tables, Means, Managing Output)

Objectives: At the end of this exercise, the participant will be able to:

- Use basic analysis functions (Frequencies, Means, Tables).
- Understand the use of routing output

Step 1: Frequencies

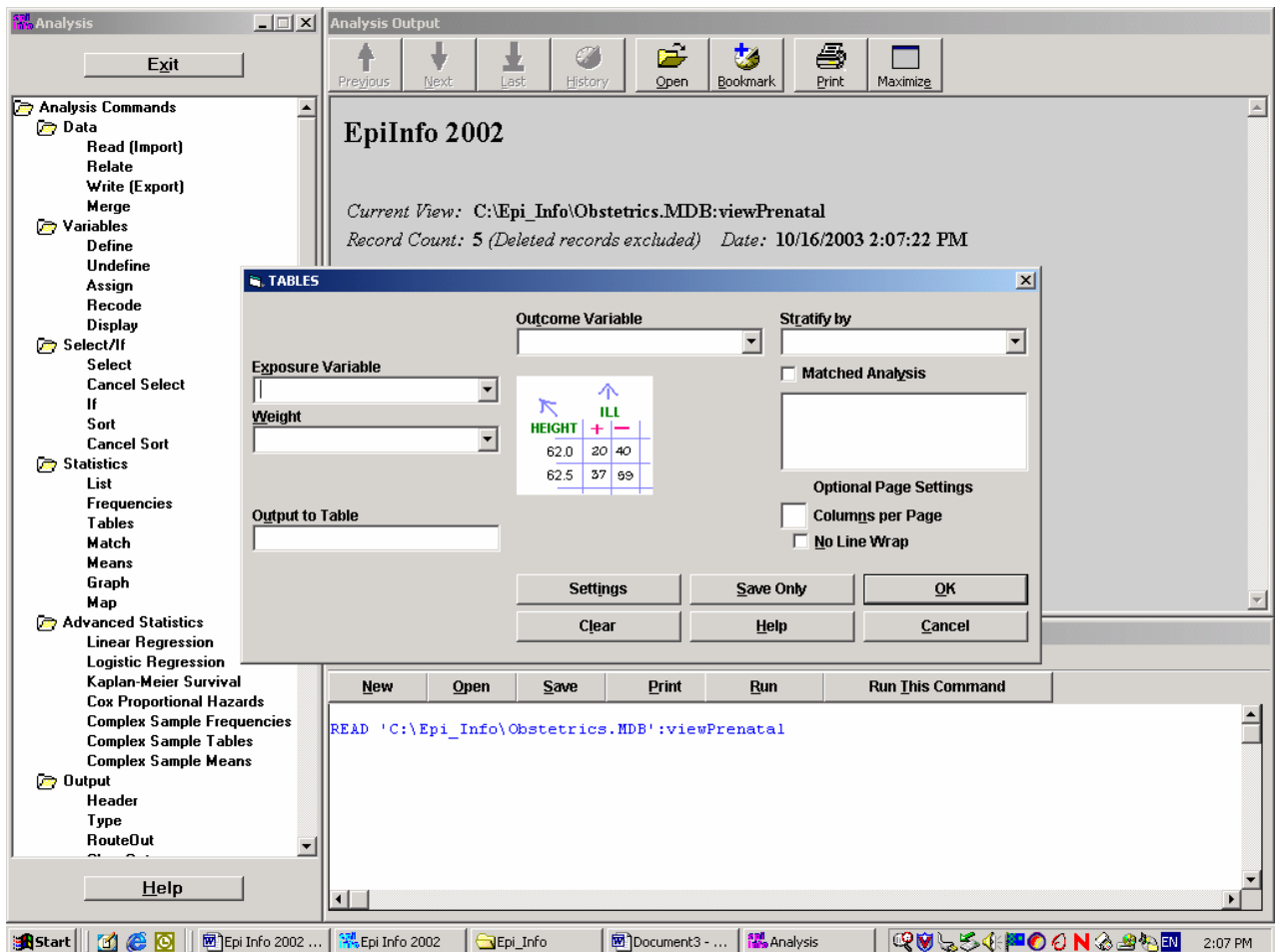
- ✓ The **Frequencies** command is used to provide frequency counts for a specified categorical variable (for example: ethnicity, gender, marital status), the percentage of the total, and the cumulative percentage.
 - a. Click **Analysis** button on main menu.
 - b. From the command tree on the left side of the screen, click on the **Frequencies** command. From the **Frequencies of** box, use the down arrow to select **Smoke**. Click **OK** and note the number of smokers and non-smokers in your dataset.
- ✓ Notice that frequencies can be calculated for more than one variable at a time or can be stratified by another variable, as well as many other options.

Step 2: Means

- ✓ The **Means** command is used to examine one continuous variable (for example, age, height, weight), while providing descriptive statistics such as mean, median, mode, and minimum/maximum values.
 - a. From the **Analysis** screen, click the **Means** command.
 - b. Select **Age** (the variable must be numeric) from the drop-down box labeled **Means of** and click on the **OK** button.
- ✓ The **Means** command can also be used to examine the relationship between a continuous variable and a categorical variable (for Analysis of Variance (ANOVA) statistics).

Step 3: Tables

- ✓ The **Tables** command is used with categorical data (variables that have no measurement scales, such as blood types, gender, race) to produce cross-tabulations.
- ✓ The relationship between two or more variables can be examined using the **Tables** command, producing 2 X 2 tables, row percentages, and other statistical outputs.
- ✓ Drop down menus are used to select the **Exposure Variable** (rows), **Outcome Variable** (columns), and a **Stratify by** variable (if desired).



- Click on the command **Tables**
- Select **Smoke** as the **Exposure (Independent) Variable** and **Alcohol** as the **Outcome (Dependent) Variable**. Click on **OK**.
- Notice the 2 X 2 cross-tabulation table produced.

Step 4: Routing output to a specific file

- ✓ **Output** commands allow you to work with the output created from analysis work.
 - ✓ **RouteOut** is a command that allows assigning a name to a specified file; otherwise, Epi Info 2002 uses a default file path (c:\Epi_Info). This default can be found by clicking on the **StoringOutput** command.
 - ✓ **CloseOut** is used to stop sending analysis output to the file specified in **RouteOut** and to continue sending to the default file.
- In the Analysis screen, **Read viewPrenatal**.
 - Click on the **RouteOut** command (under **Output** folder) in the command tree. In the **Output Filename** box, type **Risk Factors Report**.
 - Click on the **Replace any existing file** checkbox and then click on **OK**.
 - Run desired analyses (**Example: frequencies and tables for smoking and alcohol consumption**).

Step 5: *Customizing Output.*

The easiest way to manipulate output data is to cut and paste to a Word or Excel document, then make modifications.

Exercise 5: Calculating Anthropometric Measures

Objectives: At the end of this exercise, the participant will be able to:

- Calculate BMIs, Height-for-Age, and Weight-for-Age
- Save and retrieve entered data from Nutrition component

Step 1: Creating a new file and data table in Nutrition

- a. From the Epi Info main screen, click **Nutrition**.
- b. Click **File**, and then **New Table**.
- c. Enter a file name, then click **Open**.
- d. When asked if you would like to create this new file, click **Yes**.
- e. Click on the box by New Table, enter a table name, and then click **OK**.

Step 2: Enter data into the data table

- a. You can enter variables such as ID number, Last Name, First Name, Sex, Date of Measurement, Age, Height, and Weight.
- b. To calculate Body Mass Index (BMI), you must enter sex, date of measurement, birth date (or age, months is preferable), height, and weight.

Step 3: Setting Options

- a. Click on **Tools**, then **Options**
- b. This allows you to customize your data screen, choose measures and reference sources, set units, and utilize graphing functions.

Step 4: Saving records

- a. Records are saved automatically as they are entered.
- b. To be safe, you can also save your work periodically by clicking on **File** and **Save Record**.

Step 5: Retrieving existing data table

- a. Click on **File**, then **Open Table**.
- b. Enter File Name and then proceed as before.

Exercise 6: Read and Write different database formats in Analysis

Objectives: At the end of this exercise, the participant will be able to:

- Import Epi 6 and Excel data sets into Epi Info

Step 1: Opening a project from the Epi Info menu

- From the main menu, click on the **Analyze Data** button.
- Click on **Read (Import)**.

Step 2: Reading (importing) an Excel table

- Click on the **Data Formats** drop-down box and select *Excel 8.0* (or whatever version you want to import).
- Click on **DataSource (...)**, and Analysis will display a window in which you can select the Excel file you want to read. Click on your file and then click **Open**.
- Two buttons will allow you to display *Worksheets* or *Named Ranges*.
- Click on desired *Sheet* and then click **OK**. Leave the default option in the box **First Row contains field names**, and then click **OK**.
- Verify that you have access to the table by listing the dataset and calculating the **Frequency** of Sex.

Step 3: Reading (importing) an Epi 6 file

- Use the same procedure as in Step 2 to read Epi 6 .rec file, choosing Epi6 as the file format.
 - View the data using the **List** command.
 - Read (Import)** the *Obstetrics* project. Click on **All**. A data table *Address* and a view table *viewAddress* has been automatically created by Analysis.
- ✓ Note the differences between reading the different file formats. Epi Info will physically import all records inside the .REC file into a data table with the same name as the .REC file inside the current project. It will also automatically generate a view for that data table.
- Exit from Analysis.
 - From the main menu, run **MakeView** to see the new view table *Address* in the *Obstetrics* project. Exit **MakeView** when complete.

Exercise 7: Intermediate Data Management and Analysis

Objectives: At the end of this exercise, the participant will be able to:

- Define, Assign, and Recode variables
- Save, retrieve, and execute program files (.PGM)

Step 1: Define a new variable

- ✓ Newly defined variables in Analysis are used to hold the results of calculations or conditional statements.
- ✓ The value of the variable will be reset for each record as the program passes through a table. Because this is a standard variable, the last value assigned will be lost at the next **Read** statement.
- ✓ Remember to save programs so you don't have to list all of these steps every time you want to use a redefine a variable (Will be covered in Step 3 of this exercise).
 - c. To define a new variable, click on the command **Define**.
 - d. Type **LBWRisk** (LBW means low birth weight) as the name of the new variable. Then click on the **OK** button.

Step 2: Assign values to a variable based on condition (IF)

- ✓ Assign values to the defined variable **LBWRisk** based on two database variables - smoking and alcohol consumption.
 - a. To assign the new **LBWRisk** variable, click on the **If** command. From the **Available Variables** drop-down box, choose **Smoke**.
 - b. Click on the equal sign = from the code bar on the dialog box, and then click on the **"Yes"** button from the code bar. The **If** command line should look like this: **Smoke= (+)**
 - c. Click on the **OR** button.
 - d. Then from the **Available Variables** drop-down box, choose the other risk factor, **Alcohol**. Click on the equal sign = from the code bar, and then click on the **"Yes"** button from the code bar. The **If** command line should now look like this: **Smoke= (+) OR Alcohol=(+)**
 - e. Now click on the **Then** button on the dialog box. From command tree on the left, choose the **Assign** command. An **Assign** dialog box appears.
 - f. From the **Assign Variables** drop-down box, choose the newly defined variable, **LBWRisk**. Click in the = *Expression* line.
 - g. Click on the **"Yes"** button from the code bar and then click on **Add**.
 - h. Click on the **Else** button on the dialog box. From the command tree on the left, choose the **Assign** command. An **Assign** dialog box appears.
 - i. From the **Assign Variables** drop-down box, choose the newly defined variable, **LBWRisk**. Click in the = *Expression* line.
 - j. Click on the **"No"** button from the code bar and then click on **Add**. Click on **OK** to close the **IF** window.

- ✓ In the *Program Editor* section, the command you have just created from step **a.** through step **j.** should be displayed like the following text below. You may need to scroll to the bottom of the *Program Editor* to view:

```

IF Smoke= (+) OR Alcohol=(+) Then
Assign LBWRisk= (+)
ELSE
Assign LBWRisk=(-)
End

```

- k. To list the new risk variable and the risk factors, click on the **List** command.
- l. From the **Variables** drop-down box, choose **Smoke, Alcohol, and LBWRisk**. Click on **OK**.

Step 3: Saving a program file (.PGM)

- ✓ Note that each command you entered generated one or more lines of program code in the Program Editor at the bottom of the screen.
 - a. In the Program Editor, click on the **Save** button. This button will save the code written in the Program Editor in a special table in the current .MDB called Programs. A saved program can be executed with the Run PGM command or opened in the Program Editor.
 - b. In the *Program* box, type **LowBirthWeight**, then type your name in the *Author* box. Before you save, type a brief description of the training exercise in the comments box. Then click on **OK**.

Step 4: Opening an existing program

- a. In the Program Editor, click the **Open** button. This button will provide a list of programs that were previously saved.
- b. Select "LowBirthWeight" from the **Program** drop-down list. Check that your name and your comments are displayed. Click on **OK**. The program is now loaded into Program Editor.
- c. The program can be edited if desired. You can clean the saved file by deleting all the extraneous text and only saving the necessary program code.

Step 5: Running the program

- a. The program you opened is ready to be executed. Click on the **Run** button to process the program.
 - b. If you make any changes to the program, click on **Save** and save the program under the same name. Exit from Analysis.
- ✓ In some cases you may want to run only one command at a time. Place the cursor on a particular line, and then click on **Run This Command** button to execute the selected command.

Step 6: Defining a new variable to be Recoded

- In **Analysis**, open a project by clicking on **Read**.
- Click "**Change Project**", select (**Obstetrics**) then select desired view (**Prenatal**)
- Select the **Define** command from the Analysis menu tree.
- Create a variable named **AgeGroup** without any spaces. Use default standard variable type. Click **OK**. The **AgeGroup** variable has now been created.

Step 7: Recoding Ages into Age Groups

The next step is to recode the *Age* values into the new variable, *AgeGroup*.

- Select **Recode** command from the Analysis menu tree.
- Use down arrow in **From** box to select the **Age** variable.
- Use down arrow in **To** box to select **AgeGroup** variable.
- Fill in the value ranges and recoded values, using the **Enter** key to go from cell to cell. Example:

Value	To Value	Recoded Value
11	19	11-19
20	29	20-29
30	39	30-39

- ✓ The first column (**Value**) holds the lower range of the original values. The second column (**To Value**) holds the upper range. The third column (**Recoded Value**) will hold the new values of the newly defined variable, **AgeGroup**.
- e. When finished with the recoding, remember to save program code (see Steps 3-5 above) and select **OK** to exit the **Recode** dialog box. (Create a **Frequency** on the *AgeGroup* variable to make sure the values have been correctly recoded).

Step 8: Saving the changes into a new data table

- ✓ The **Define** and **Recode** commands create temporary variables and values. Closing Analysis and reading another table or database will erase any defined or recoded variables.
- ✓ To keep the new variables that were defined (without having to remember to run the saved program commands every time), use the **Write (Export)** command.
 - Select the **Write (Export)** command from the Analysis menu tree.
 - Select **Replace** under **Output Mode** if you want to write over previous files or select **Append** if you want to add new data onto an old file. Locate **Obstetrics** by clicking on the (...) button near **File Name**. At the line **Data Table**, create a new table called, **Obstetrics2**. Make sure **All** variables are selected.
 - Read (Import)** **Obstetrics2** to verify that the table was saved properly. Remember to click on the **All** button to see the new table (a new View has not been created for it yet). **List** the table and verify that the data table was saved properly.

IV. References

Alperin, M. and Miner, K, Using Epi Info 2002: A Step-by-Step Guide, 2003 (This is an excellent resource).

CDC Epidemiology Program Office Epi Info website: www.cdc.gov/epiinfo/about.htm

CDC, Epi Info 2000 Manual (The updated 2002 manual to be released a few months from this printing)

CDC, Epi Info 2002 Training Manual, March 10-13, 2003.