

**Manual for the Crosstabulation of CBMS Data
using CPro Crosstabs**

Prepared by the
CBMS Network Coordinating Team of the
Angelo King Institute for Economic and Business Studies

This work was carried out by the PEP-CBMS Network Coordinating Team with the financial support of the Government of Canada provided through the International Development Research Centre (IDRC) and the Canadian International Development Agency (CIDA)

Manual for the Crosstabulation of CBMS Data using CPro Crosstabs

I. Introduction

A. About CBMS-CPro

The Census and Survey Processing System (CPro) is a software package for entering, editing, tabulating, and disseminating data from censuses and surveys. It is designed to fulfill the census and survey data processing needs of data-producing organizations worldwide. This state-of-the-art software combines the best technical features of earlier software with a user-friendly interface to enhance ability to collect, analyze, and disseminate data. CPro is developed by MEASURE partners, the U.S. Bureau of the Census, ORC Macro's MEASURE DHS+ project, and SerPro S.A. CPro runs under Windows 98se, Me, NT 4.0, 2000, XP, Vista or 7. It is a public domain product, so it can be used and distributed at no cost.

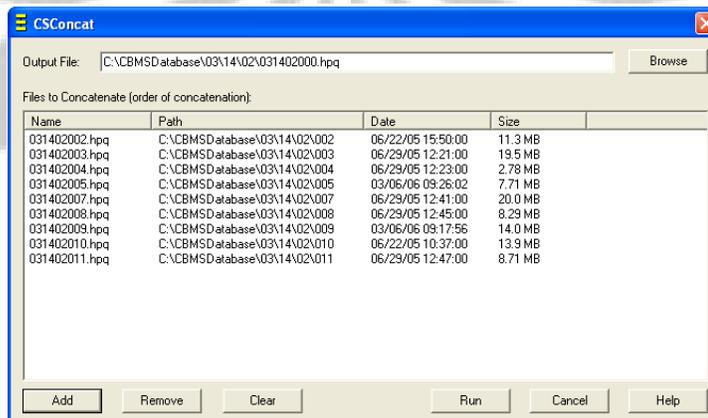
B. Installation

The CBMS Network Coordinating Team had customized the Crosstabulation application of the CPro for the use of the local government units. The HPQ and BPQ Crosstab are installed simultaneous with the installation of the CBMS-StatSim.

II. Preparing Encoded Data for Crosstabulation

Completely encoded barangay-level data is ready to be used for crosstabs. However, if all barangays in a municipality had already been completely encoded, all the barangay files can also be concatenated so that crosstabulation can be done at the this level.

The following figure shows an example of concatenation of barangays to form a text file of a municipality, for example: data for the Municipality of Balagtas:



Note in the figure above shows that the output file path is:

This indicates that a text file (031402000.hpq) will be created containing the data from the "Files to concatenate" window.

Note that the path of inputs and output are fixed, that is, each of the final encoded barangays is inputted and the output is now the text file of the municipality in the municipal directory. Thus in general, merging the barangays of a certain municipality shall have inputs:

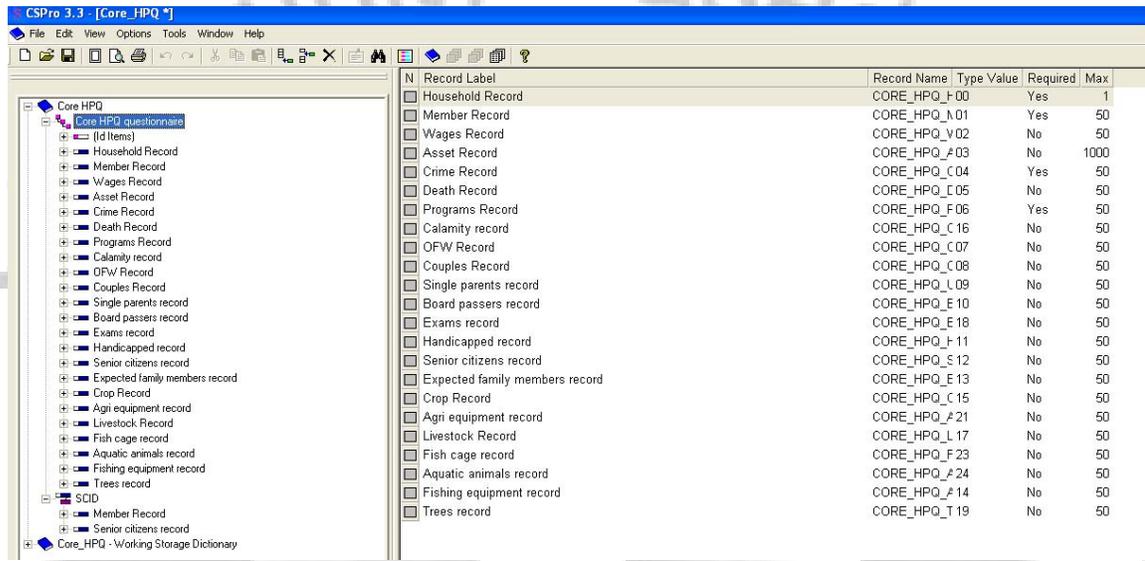
C:\CBMSDatabase\<Regn>\<Prov>\<Mun>\<Brgy>\<Brgy.txt>

And the output will be: **C:\CBMSDatabase\<Regn>\<Prov>\<Mun>\<Mun.txt>**

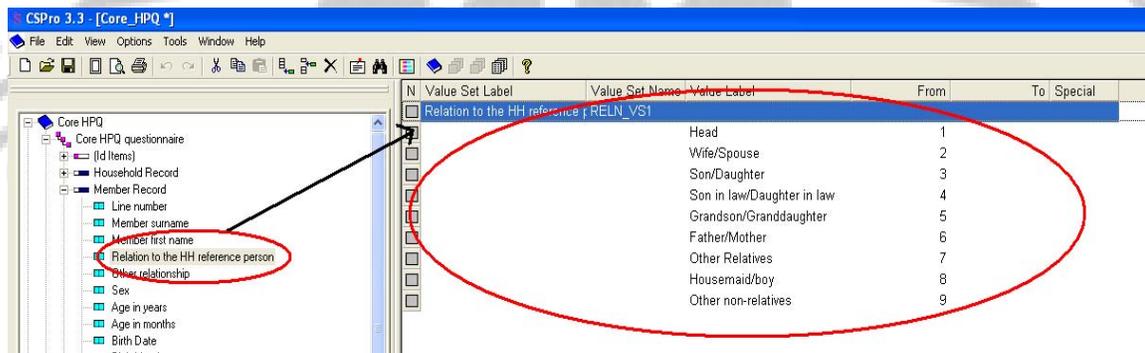
III. Data Dictionary

A Data Dictionary describes the overall organization of a data file. It gives a description of how data are stored in a file. Each dictionary will allow you to give text labels for all levels, records, items, and value sets in the file. It also allows you to describe the organization of each record in the file and the characteristics of each item in the record: names, position in the data record, type of data, length, number of decimal places, valid values, and other documentation.

Data Dictionary Screen Lay-out for the CSPro-based encoding system



The data dictionary screen is divided into 2 windows. The screen on the left displays the dictionary tree. The screen on the right displays detailed information for the highlighted object in the left-hand screen. For example, if in the left-hand screen the focus is on the first line (**Core HPQ**), the right-hand screen will display information about the dictionary file as a whole. If the focus in the left-hand screen is moved to the second line questionnaire or case level, the right-hand screen will display information about the questionnaire or case, which is the basic element of the file.



As the focus is moved down the dictionary tree, the right-hand screen changes to reflect the different items of interest at each successive level.

IV. Introduction to Tabulation

CSPro has a tabulation application which allows you to tabulate data quickly and easily, producing basic frequency distributions and cross-tabulations from one or more data files. The tabulation application has many capabilities. One of these is Cross Tabulation.

A. What is Cross tabulation?

Cross-tabulation is taking a minimum of two data items and tabulating the results of one data item against the other data item. There may be one independent and one dependent variable in each dimension [row and column]. Cross tabulation gives a basic picture of how two variables inter-relate, and can help in determining patterns of interaction. Like frequency distributions, cross-tabulation results may be displayed as numeric values [default], as percentages, or as both numbers and percentages.

B. Parts of a Table

1. What is a table?

A table represents relationships between data set. It is a set of values that is organized using a model of vertical columns and horizontal rows. A table has a specified number of columns, but can have any number of rows. There are five basic parts of a table.

HEADING			
Table 20. Barangay by Sex		COLUMN	
Barangay	Total	Male	Female
Total	11564	6080	5484
Antipuluan	4505	2312	2193
Aramaywan	3499	1868	1631
Batang-batang	1947	1056	891
San Isidro (Bato-bato)	1613	844	769
Burirao	-	-	-

ROW (points to the 'San Isidro (Bato-bato)' row)
CELL (points to the '844' value in the 'Male' column for 'San Isidro (Bato-bato)')

FOOTNOTE

Note: no data for Burirao
Source: CBMS survey 2004

- Heading – is the first line in the table. This is also called as the title of the table. It describes the contents of the table.
- Column – shows the column variables and describes the contents of a column of data.
- Row – shows the row variables and describes the contents of a row of data.
- Cell – is one grouping within a table. Cells are grouped horizontally (rows of cells) and vertically (columns of cells). Usually information on the top header of a table and side header will "meet" in the middle at a particular cell with information regarding the two headers it is collinear with.
- Footnote – note of text placed at the bottom of a table, and usually contains the data source.

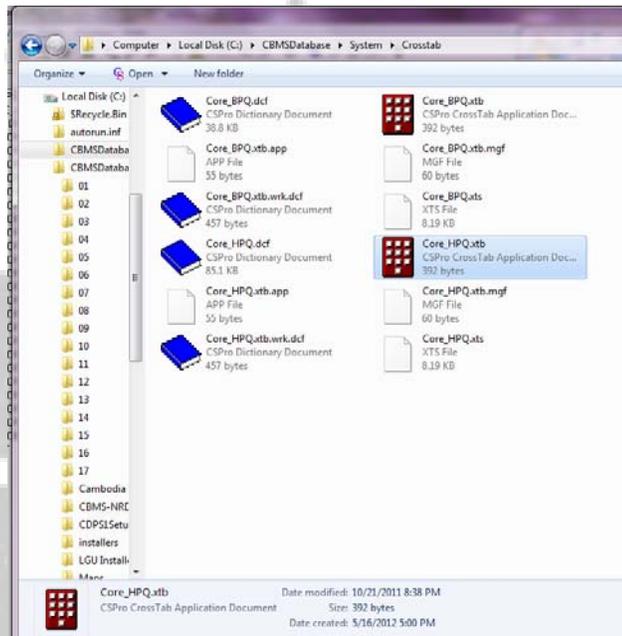
V. CPro Crosstabulation

A. Opening the CPro Crosstabulation

There are two ways to open the CPro Crosstabs.

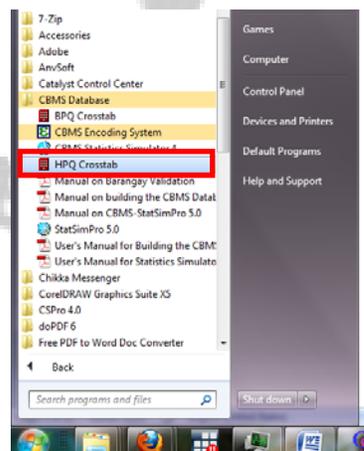
1. First Method

- a. Under the C directory find the folder with file name *CBMSDatabase*
- b. Look for the *System* folder
- c. Open the *Crosstab* Folder
- d. Select Click *Core_HPQ.xtb*



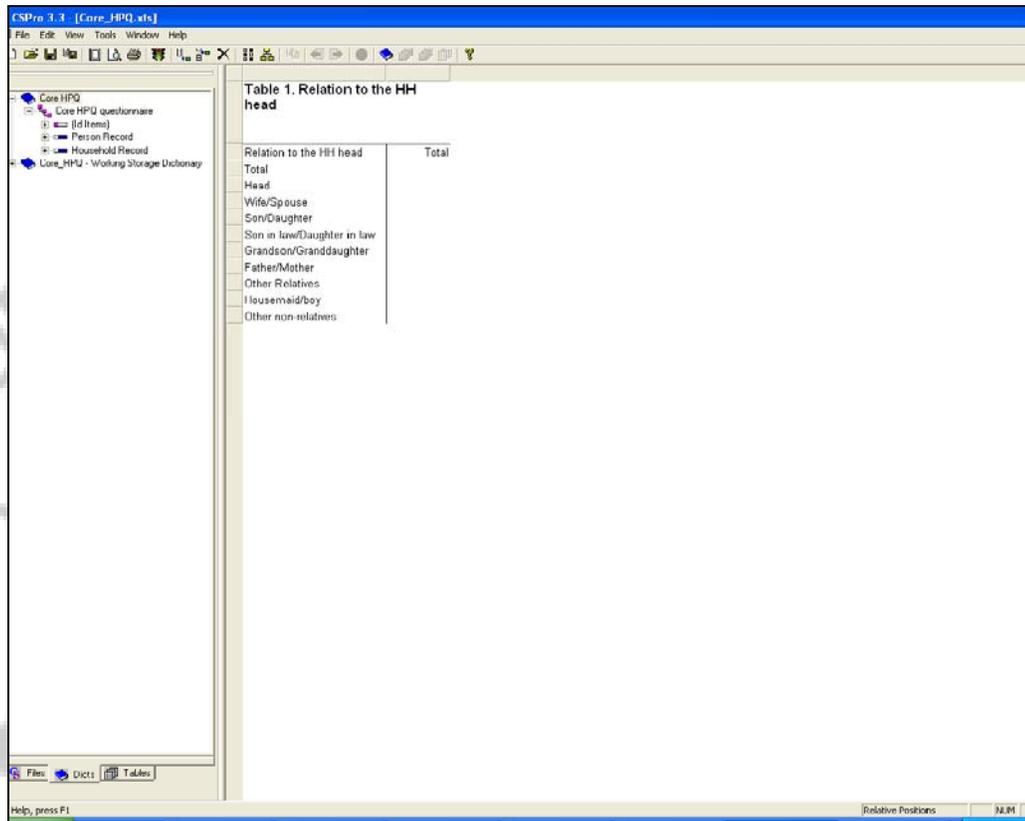
2. Second Method

- a. Click *Start*
- b. Click *All Programs*
- c. Search for *CBMSDatabase* and Click it
- d. Select *HPQ Crosstab*

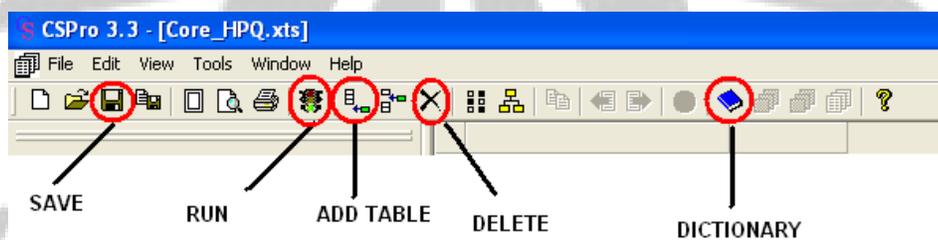


B. Cross tabulation Interface

After opening, a window such as below will appear.



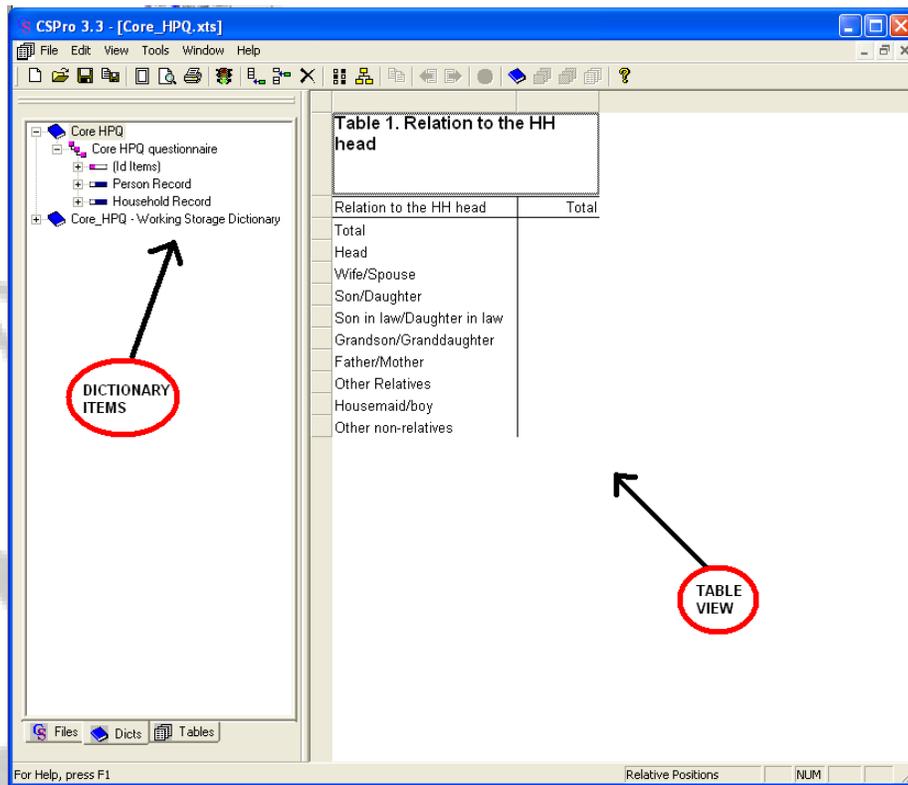
C. Getting to Know Cross tabulation



In the upper part of the crosstab window or toolbar, there are shortcut icons that can be used. Some commonly used shortcut icons:

1.  - Click this to *Save* the created table
2.  - After creating any table click this *RUN* icon to process the data needed for the table. Values will then be shown in the table.
3.  - To *add/create* a new table
4.  - To *Delete* a table
5.  - To view the contents of *Data Dictionary*, which defines the data or variable.

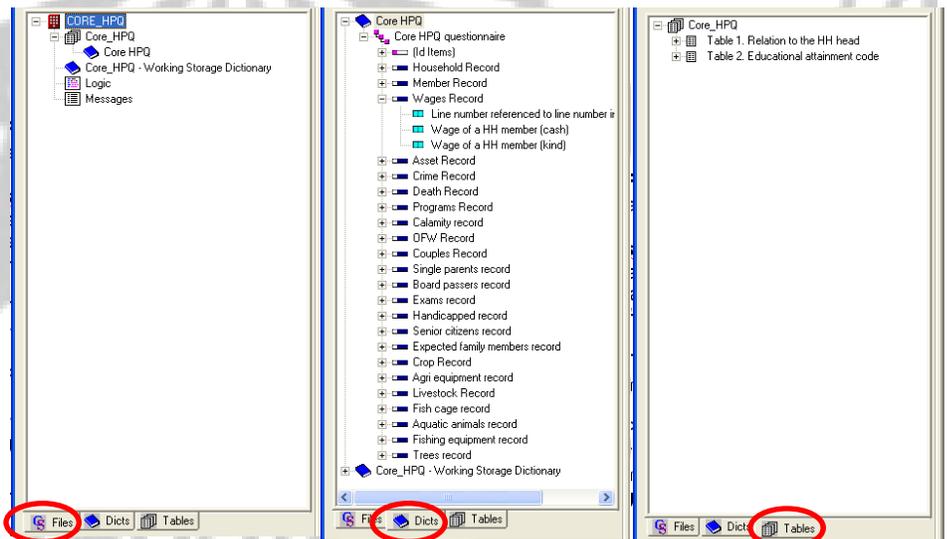
The Crosstabs window is divided into two parts. The left hand display contains three tabs at the bottom -Files, Dicts, and Tables.



The **Files tab** displays the list of files that are being accessed or used.

Dicts tab shows the variables that are available. There are many records/tables available. The ID items define the regions, municipality, barangays and purok. Member record contains a list of all variables about the household members such as age, sex, grade/yr level and others.

When "+" is selected for each corresponding category, a list of fields will be shown. Field  represents the individual variable/item. If the item has more than one value set, then the field will show the symbol "+" indicating that the item can be expanded. Value sets are always designated with the symbol  called value set icon.



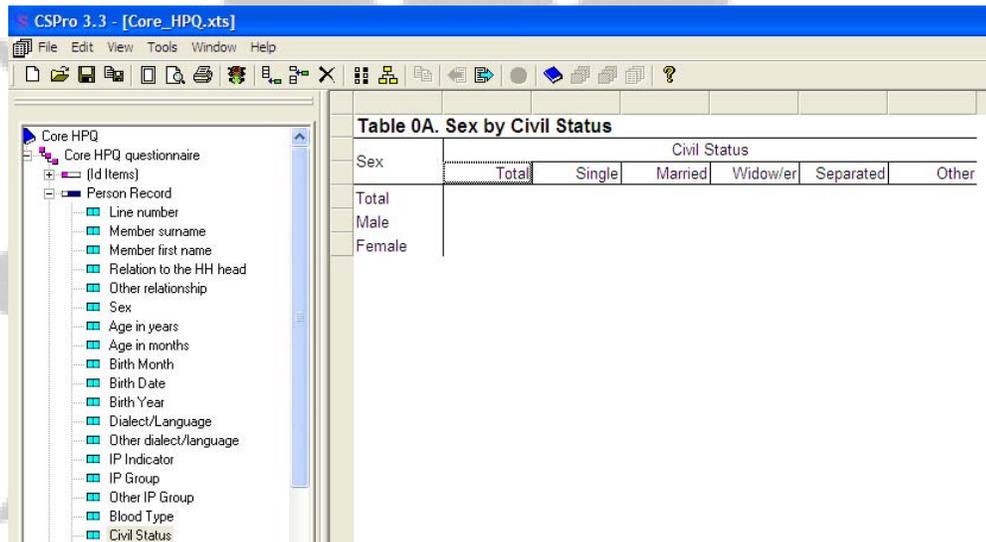
The last tab called **Tables** displays a list of tables created.

VI. Outputs of Cross tabulation

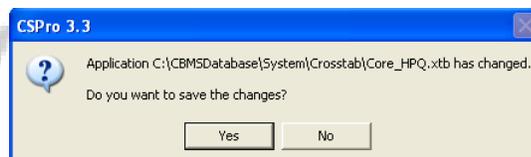
A. Basic Crosstab

To create a table, do the following:

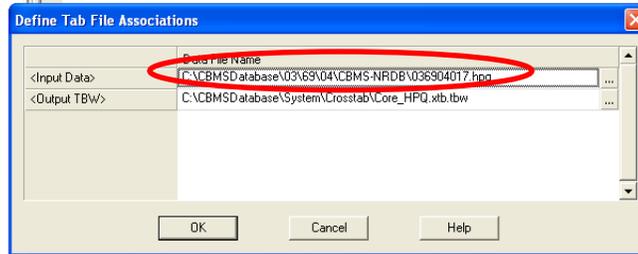
1. Click the shortcut icon  to add a table.
2. Select the Dictionary [Dicts] tab to make the dictionary file structure visible.
3. Expand the tree until the item(s) to be used for row or column variable(s) appears on the tree.
4. Drag the desired variable to the right and drop it in the table view. Where we drop it on the table will determine whether it is used as a row or column variable. Imagine a diagonal line connecting the top-left corner of the table window to the bottom-right corner, dictionary items dropped below and to the left of the line will become row variables and items dropped above and to the right of the line will become row variables.
5. Repeat Step d for a total of two items (value sets) per row and two items per column.



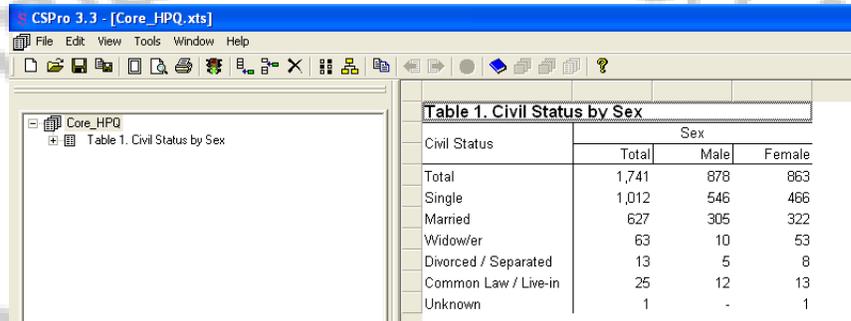
6. To delete a row/column variable, left-click on one of its value names, then drag and drop it back on the dictionary tree.
7. Click the icon  to run the data.
8. A dialog box such as below will appear. Click Yes.



After clicking yes, another window will appear. In the Input data, click the three dots found in the end, and select the text file that contains the barangay, municipal or province data that will be used in the table. Then Select OK.



Below is an example of created table for Civil Status and Sex.



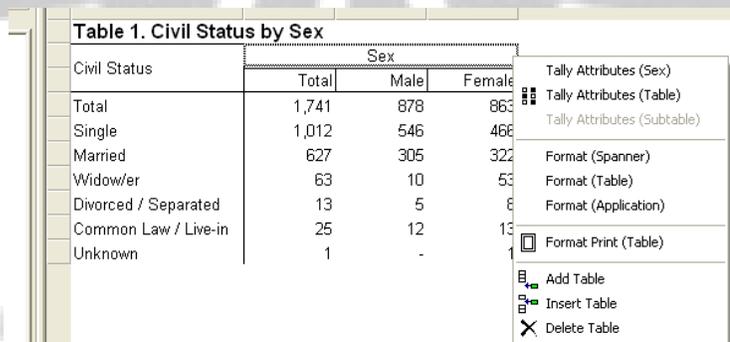
Civil Status	Sex		
	Total	Male	Female
Total	1,741	878	863
Single	1,012	546	466
Married	627	305	322
Widow/er	63	10	53
Divorced / Separated	13	5	8
Common Law / Live-in	25	12	13
Unknown	1	-	1

B. Percents

Aside from numbers or magnitude, cross tabulation enables us to show the distribution of values for an item as a percentage of either row or column totals, or as a percentage of the table total.

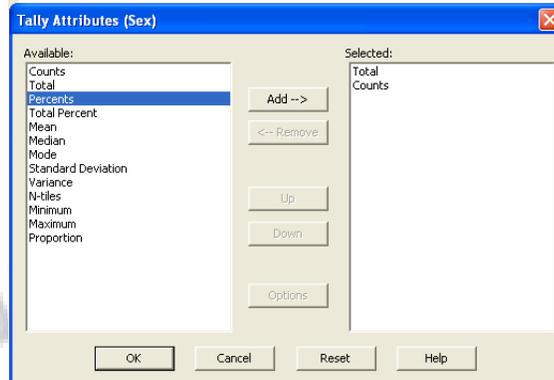
There are optional percentages available for each value set in a table. Tally attributes is used to create percentages. To bring up the Tally Attributes

1. Right-click on the variable in the table and choose Tally Attributes (<variable>). In this example, select Tally Attributes (Sex).

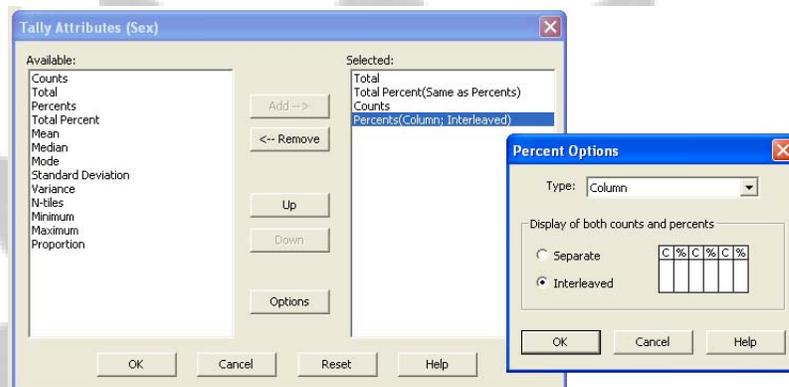


Civil Status	Sex		
	Total	Male	Female
Total	1,741	878	863
Single	1,012	546	466
Married	627	305	322
Widow/er	63	10	53
Divorced / Separated	13	5	8
Common Law / Live-in	25	12	13
Unknown	1	-	1

2. A window shown below will appear.



3. Select "Percents" in the list of available calculations on the left-hand side of the dialog and click the "Add" button. Then highlight Percents (Column; Interleaved) and click the button "Options".



4. There are three types of percentages users can opt:
 - i. **Total** – Percents are based on total cell for the table ($\% \text{ cell} = (\text{associated cell value} * 100) / \text{total of all the cells in the table}$).
 - ii. **Row** – Percents are based on total for the row ($\% \text{ cell} = (\text{associated cell value} * 100) / \text{total of all the cells in the row}$).
 - iii. **Column** – Percents are based on total for the column ($\% \text{ cell} = (\text{associated cell value} * 100) / \text{total of all the cells in column}$).
5. Interleaved/Separate

By default, percents and counts are interleaved; that is, for each value in the value set, the frequency for that value is listed in the column/row directly preceding or following the percent for that value. Percents may also be separate, in which case all the frequencies are grouped together and all the percents are grouped together.

Sex					
Total	Percent	Male	Percent	Female	Percent

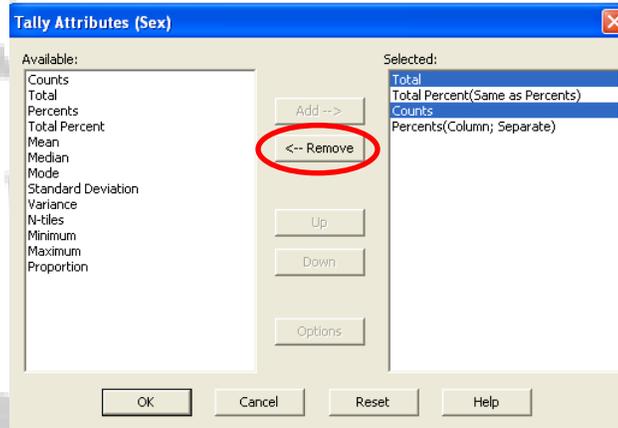
Interleaved Percents

Sex					
Total	Male	Female	% Total	% Male	% Female

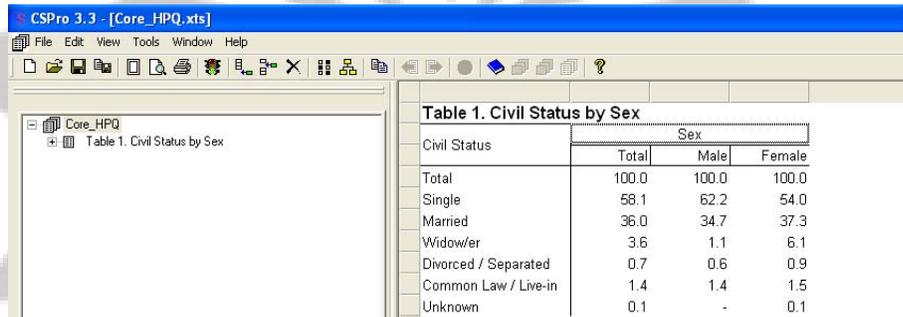
Separate Percents

For percents to be interleaved, the percents must directly follow or directly precede the counts in the list of selected calculations in the tally attributes dialog. The position of the percents can be changed relative to the other selected calculations (including the counts) by selecting the percents in the list on the right-hand side of the dialog and clicking the "Up" or "Down" buttons.

- To show percents only (percents without counts), add the percents as described above and set the options to "Separate". Then select the counts in the list of selected calculations on the right-hand side of the dialog and click the "Remove" button to delete the counts. Do the same to remove the "Total". Click OK.



- Repeat steps VI. A. 7 and 8 to generate the table below.



Civil Status	Sex		
	Total	Male	Female
Total	100.0	100.0	100.0
Single	58.1	62.2	54.0
Married	36.0	34.7	37.3
Widow/er	3.6	1.1	6.1
Divorced / Separated	0.7	0.6	0.9
Common Law / Live-in	1.4	1.4	1.5
Unknown	0.1	-	0.1

C. Tabulate Values and/or Weights

Cross tabulation also allows the use of a data item or a constant value as a weighting factor. This is particularly useful in the case of a survey, where the weight assigned to each case or observation in the sample must be taken into account in order to produce numbers representative of the whole. If no weight value is specified, weight is assumed to be "1". Like weights, values are numeric data items or constants with or without decimal positions. When a value is specified for tabulation, the effect is that of cumulative addition of the specified value into the cell at the intersection of the row and column coordinates. If value item is blank, a value of 1 will be added into the appropriate cell during tabulation. If both value and weight are specified for a given table, the specified value is first multiplied by the specified weight and the product of this multiplication is then added to the cell in question. Select the universe of tabulation.

D. Summary Statistics

Cross tabulation can also include summary statistics in the tables. Most of these give information about the distribution of values for the data item.

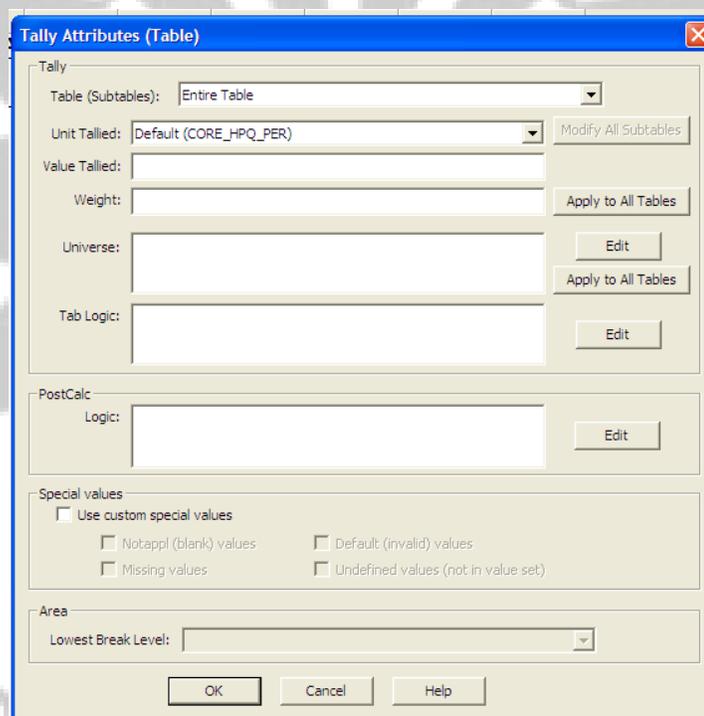
To add a summary statistics, right click on the variable in the table. Choose Tally Attributes (<variable>) and select the desired measure. The choices for statistical measures are:

- **Mean** – average value of observations.
- **Median** - middle value (half of the observations are above this value and half below).
- **Mode** – for categorical variables this is the most frequently observed value. For grouped variables, this is the lower limit of the modal class or the range in the value set in which the most observations lie.
- **Std Deviation** – Standard Deviation, a measure of how clustered the observations are around the mean or the square root of the variance.
- **Variance** – another measure of dispersion.
- **N-tiles** - values that divide the values into N groups each of which contain 1/N of the total observations (N = 2 is equivalent to median).
- **Minimum** – smallest value found in all observations.
- **Maximum** – largest value found in all observations.
- **Proportion** – show counts for certain values in the value set as a fraction or percentage of the total for the variable.

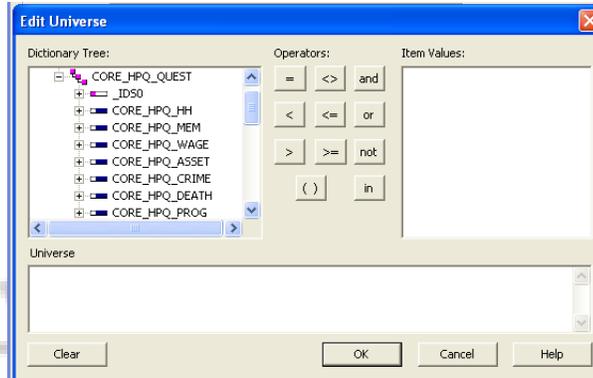
E. Universe

The "universe" specification acts as a filter, as the tables produced use only a subset of the data file's records. It limits the scope of the data being outputted. Therefore, values in the table may be lower than they would be with no universe specified, since the universe restricts the data available for tabulation. To do this, follow the steps below:

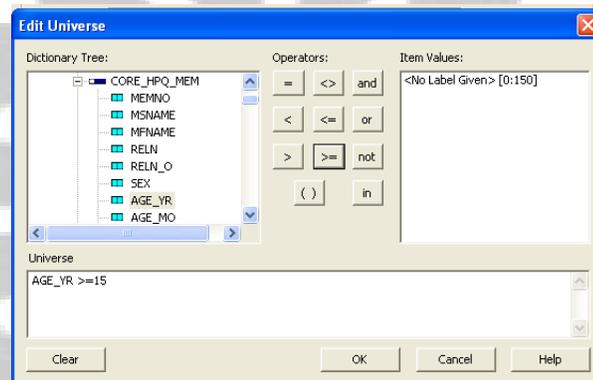
1. Right click on the variable in the table. Choose Tally Attributes (<table>). A dialog box will appear.



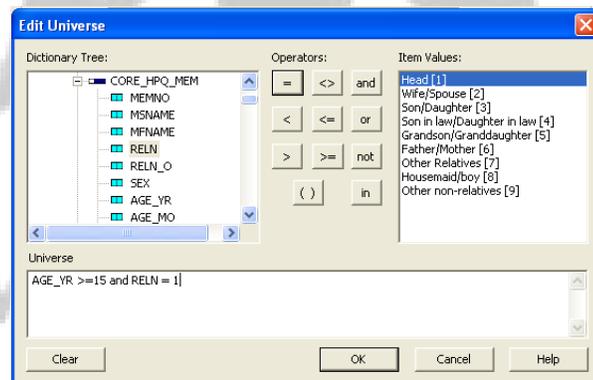
2. Click edit in the corresponding universe field. Another dialog box to edit the universe will appear.



3. Expand the Dictionary tree on the left hand side of the window until the desired variable to be restricted is visible.
4. Double click the variable. Automatically, the value for the selected item will be shown in the item value. Use operators to define the universe. Example, let us use restrict the table to show only data for ages 15 and above. (AGE_YR>=15). Note: Refer to Annex 1 of the User's Manual for CBMS Statistics Simulator (StatSimSGE version 4.0) for the list and meanings of the variables.

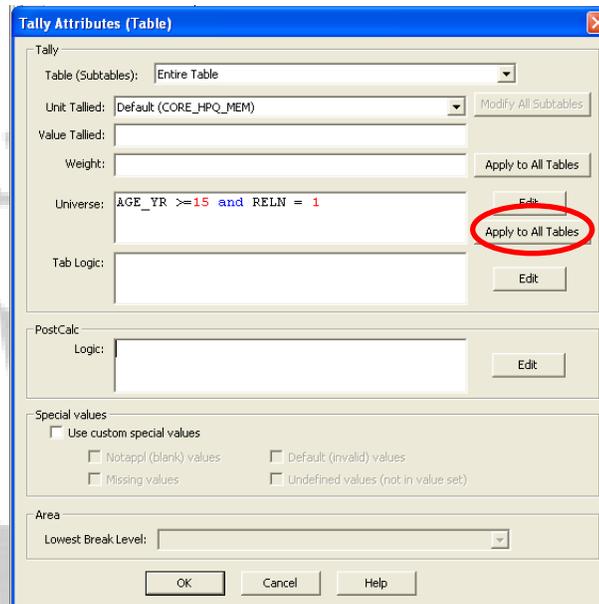


5. To enter several conditions, use and / or. Add parentheses to modify the order of evaluation of the conditions. Adding to the previous example, aside from age, the population was also restricted to contain data only for all household heads (RELN=1)

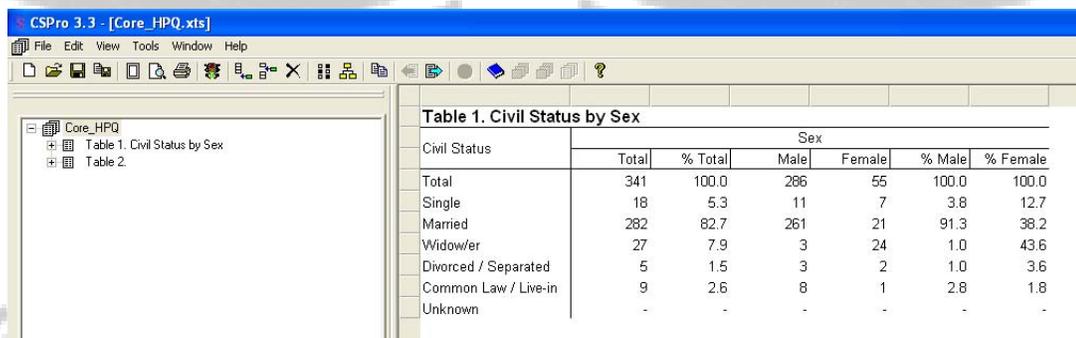


6. To delete the universe simply erase the contents of the universe condition.
7. Press OK when the universe condition is complete.

- To apply a universe to all the tables, click the Apply All button to the right of the Universe box in the Tally Attributes dialog box. Please note that if there is some type of error either in the typed or generated condition, the system will display an "Invalid Universe Syntax" message and the problem must be corrected before the universe can be accepted.



Here is the resulting table after applying the restriction to the universe. Note that from the list earlier in VI.A.8. in page 9, from 1,741 as the total population, there are 341 persons who are aged 15 and above and household head.

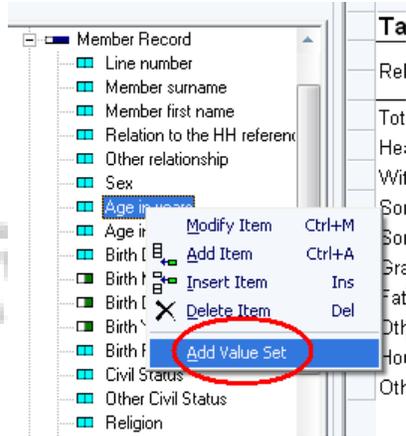


Civil Status	Sex					
	Total	% Total	Male	Female	% Male	% Female
Total	341	100.0	286	55	100.0	100.0
Single	18	5.3	11	7	3.8	12.7
Married	282	82.7	261	21	91.3	38.2
Widow/er	27	7.9	3	24	1.0	43.6
Divorced / Separated	5	1.5	3	2	1.0	3.6
Common Law / Live-in	9	2.6	8	1	2.8	1.8
Unknown	-	-	-	-	-	-

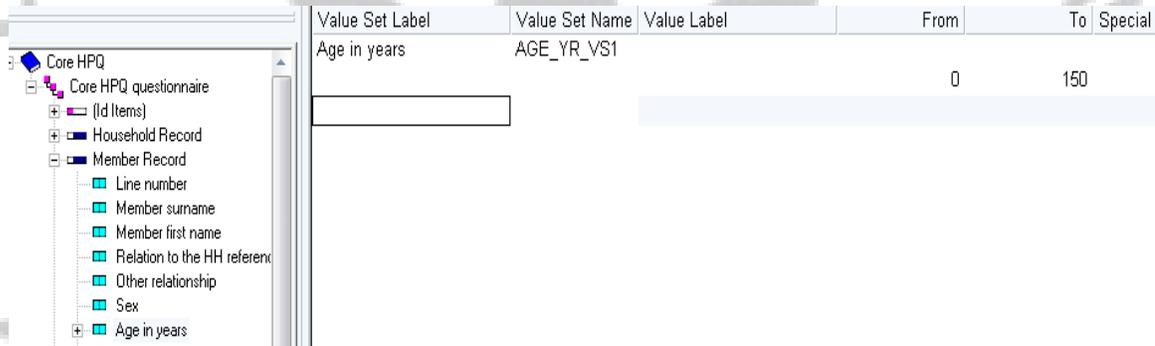
F. Value Sets

Value sets let you specify one or more group of values for a data item or subitem. The resulting tables will contain row and column labels that correspond to the value labels (or numeric distributions, if no value label is present). For example, we want to classify ages of the population such as 0-5, 6-12, 13-16, 17 and above. Here are the steps:

- Expand the dictionary tree. Select the item or variable we wish to add a value set. In our example, Age_yr.
- Right click and select "Add Value Set", or select "Modify Value Set" to modify any existing value set.



3. Automatically, it creates a new value set name shown in the right hand side. It also requests some value set properties such as Value Set Label, Value Label, From, To and Special



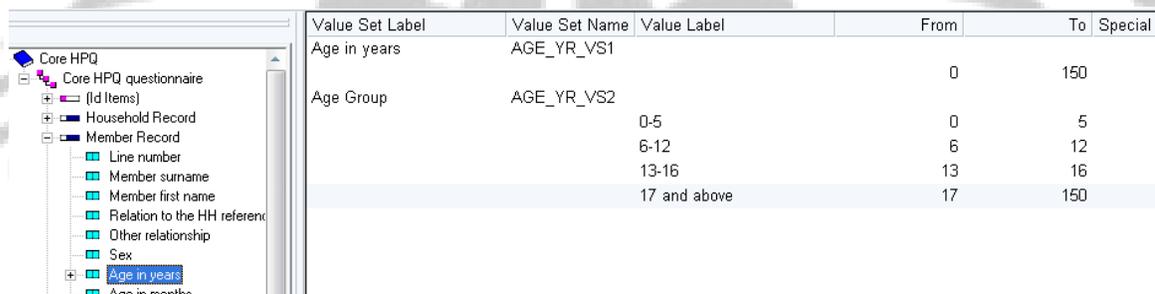
Value Set Label: A text label that describes an item or variable. This cannot be left blank.

Value Set Name: The name of this value for use in the CSpPro language procedures.

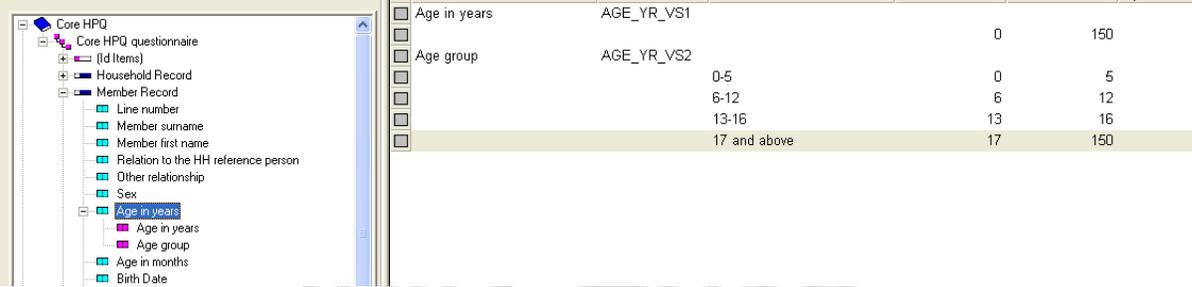
Value Label: A text label that appears in the table.

From: The smallest value in the value set.

To: The largest value in the value set.

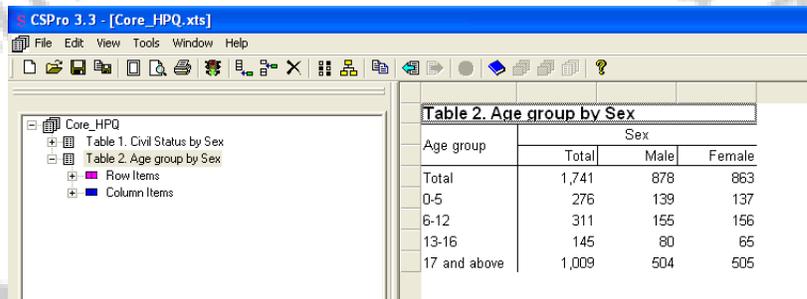


4. After Completing the value properties requested, press the **Esc** key. A single value set can contain one or more values. The value will always be added to the end of the value set listings. Use also the **Esc** key when the value is added to a wrong place. Note that you will have two value sets for age in years: age in years and age group.



Age in years	AGE_YR_VS1		
		0	150
Age group	AGE_YR_VS2		
	0-5	0	5
	6-12	6	12
	13-16	13	16
	17 and above	17	150

5. Drag age group and sex in the table. Here is the resulting Crosstab.



Age group	Sex		
	Total	Male	Female
Total	1,741	878	863
0-5	276	139	137
6-12	311	155	156
13-16	145	80	65
17 and above	1,009	504	505

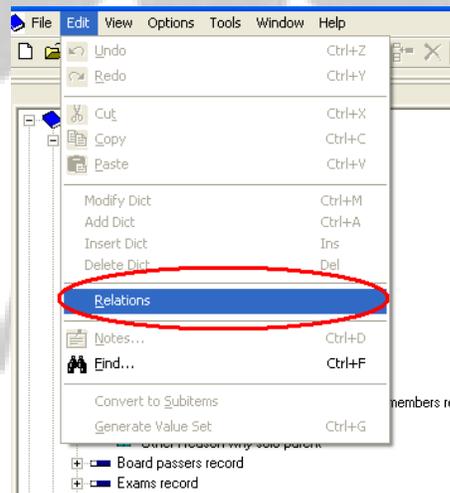
G. Relations

Relations provide a way of linking one multiple record or item to one or more multiple records or items. Once the items on different records are linked through a relation, tallies are done just as if all the items belonged to one record. Relations work much like database joins.

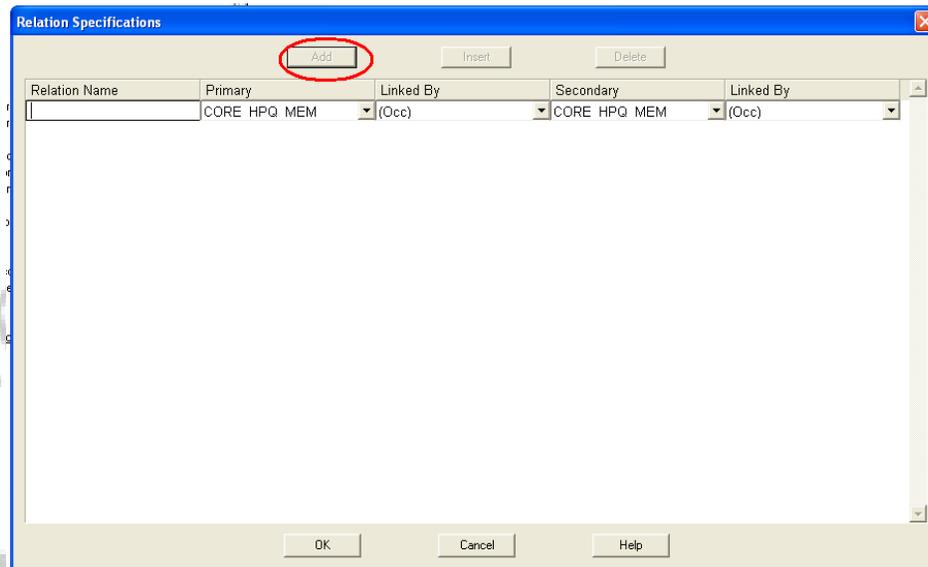
Relations have one primary multiple record or item. Each instance of the primary element in a case is processed one at a time. A relation has one or more secondary records or items. The corresponding secondary elements are linked to the primary element during processing.

For example, we want to cross tabulate Senior citizens with ID by Sex. We should first add a relation to link the Senior Citizens record with the Member Record. To do this, here are the steps:

1. In the menu toolbar click *Edit* and select *Relations*.

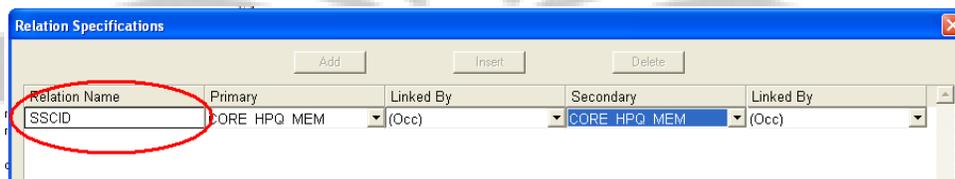


2. A dialog box will then appear. Click the *Add* button to add a relation at the end of the list or press the *Insert* button to insert a relation at the current highlighted line. Click *Delete* button to delete the highlighted line.



Relation Name	Primary	Linked By	Secondary	Linked By
	CORE HPQ MEM	(Occ)	CORE HPQ MEM	(Occ)

3. Notice that there are five Relation properties to be filled. The first relation property is the relation name. **Relation name** is arbitrary. Choose a relation name that describes the new relation. For example, we will use SSID as the relation name to denote Senior Citizens ID.



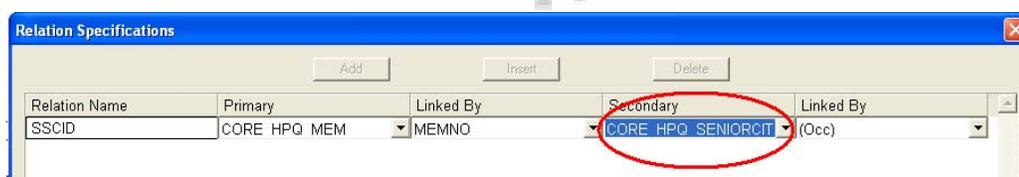
Relation Name	Primary	Linked By	Secondary	Linked By
SSCID	CORE HPQ MEM	(Occ)	CORE HPQ MEM	(Occ)

4. The **primary** property is the name of the record or item to be linked to the secondary item. For our example, we will use Core HPQ_mem.
5. The **Linked by** is the name of an item that links the primary and secondary. If the primary is a record then this is the name of an item within the primary record. We will choose Memno as example. Since the number on each record (memno and seniorcit_line) that links the two records which identifies a single person is the same on both records.



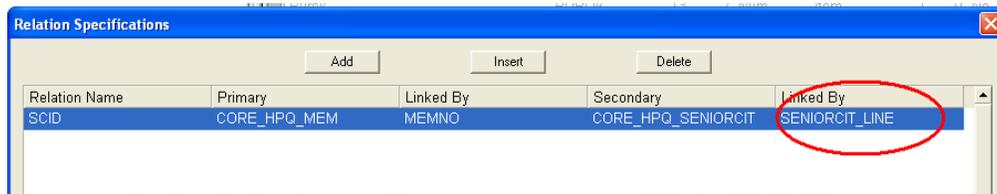
Relation Name	Primary	Linked By	Secondary	Linked By
SSCID	CORE HPQ MEM	MEMNO	CORE HPQ MEM	(Occ)

6. Items within the **secondary** are linked to item in the primary. The secondary cannot be the same as the primary. In the secondary field, choose core HPQ_seniorcit

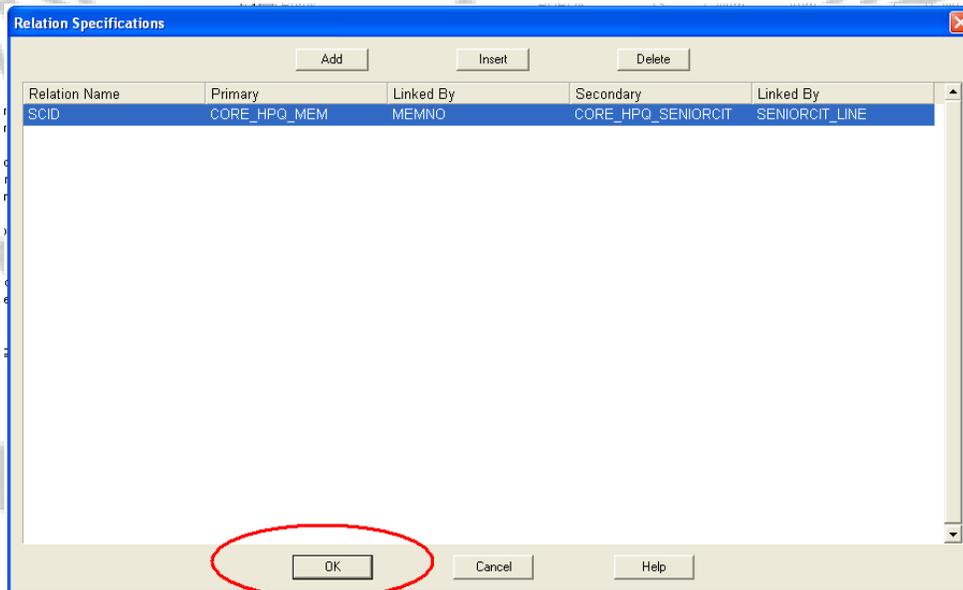


Relation Name	Primary	Linked By	Secondary	Linked By
SSCID	CORE HPQ MEM	MEMNO	CORE HPQ SENIORCIT	(Occ)

7. As mentioned in Step e, the **linked by** is the name of the item that links the two records. If the secondary is a record, then this is the name of an item within the secondary record. Choose the item Seniorcit_line for our example.



8. Click the OK button if all the relation properties are completely filled-up.

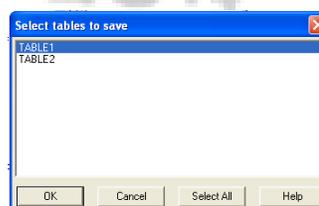


9. You are now ready to cross tabulate Household member with senior citizen ID by Sex. Please note that the added item or variable is usually at the end of the dictionary tree under your chosen relation name.

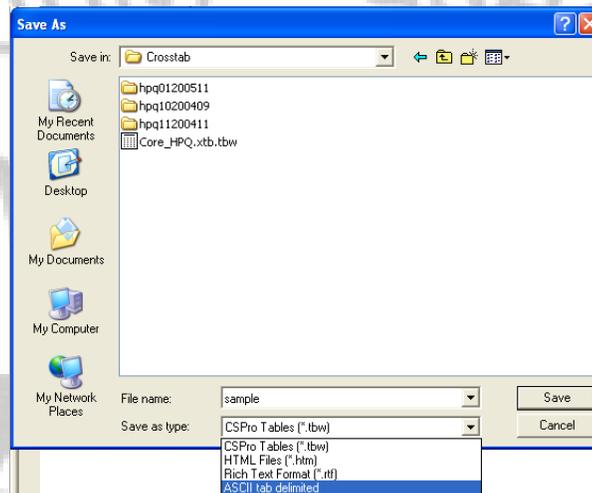
VII. Saving and transferring Crosstab outputs

A. Save tabulations in several formats.

1. To save all tables in the Crosstab application,
 - a. Click  on the toolbar; or from the File menu select Save Tables.
 - b. If there is only one table in the Cross Tabulation application, you will go directly to the Save As dialog.
 - c. If your Cross Tabulation application has multiple tables defined within it, a Select tables to save dialog box listing the tables is displayed.



- d. Use the Select All button if all tables are needed. Otherwise, select the individual table(s) that you would like to save in a single Table Viewer file. (Multiple tables are selected in the usual manner with the Shift and/or Ctrl keys.)
- e. After selection press OK.
- f. In the "Save As" dialog box use the drop down "Save as type" menu to select the type of file other than .tbw.



- g. Enter the name of the file to be created or browse to select the name of the file to be replaced.

Note: ONLY ONE table at a time can be in Rich Text (.rtf) or HTML (.htm) format. ASCII tab delimited Format supports saving multiple tables in a single file.

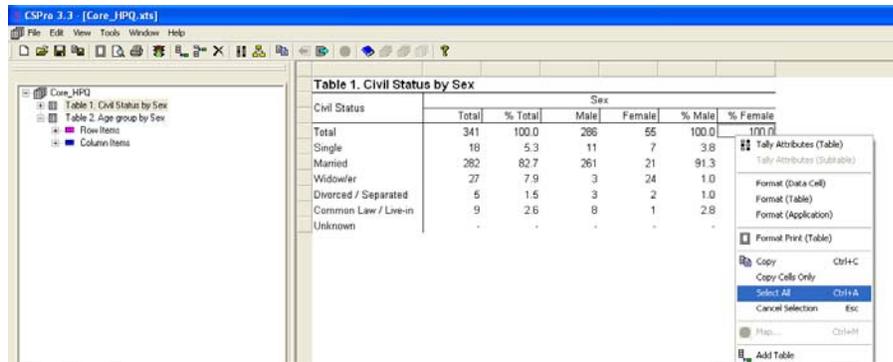
The crosstabs created can be saved in several formats:

- **Rich Text Format (.rtf):** Can save tables in rich text format so they can be used later by word processors such as Word or WordPerfect. The tables with *.rtf extension when opened in word processor will appear in the word processor's table format.
- **HTML files (.htm):** Can save tables as HTML files so they can be later incorporated into Internet applications in table format.
- **ASCII tab-delimited (.other):** Can save tables in other formats so they can be used later by a spreadsheet such as *Excel*. We can open the file in spreadsheet, and the table will appear as a matrix of cells with columns lined up.

B. Copy tables to spreadsheets or word-processing documents.

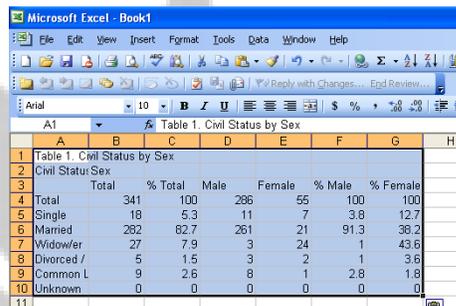
In crosstabs, the entire table can be copied to the clipboard. To copy or paste a table:

1. from the table itself, right-click anywhere over the cells of a table. A pop-up dialog box will appear. Select "Select All".



Civil Status	Sex					
	Total	% Total	Male	Female	% Male	% Female
Total	341	100.0	286	55	100.0	100.0
Single	18	5.3	11	7	3.8	3.8
Married	282	82.7	261	21	91.3	91.3
Widower	27	7.9	3	24	1.0	1.0
Divorced / Separated	5	1.5	3	2	1.0	1.0
Common Law / Live-in	9	2.6	8	1	2.8	2.8
Unknown	-	-	-	-	-	-

2. Click copy and paste into word or excel files.



	A	B	C	D	E	F	G	H
1	Table 1. Civil Status by Sex							
2	Civil Status: Sex							
3		Total	% Total	Male	Female	% Male	% Female	
4	Total	341	100	286	55	100	100	
5	Single	18	5.3	11	7	3.8	12.7	
6	Married	282	82.7	261	21	91.3	36.2	
7	Widower	27	7.9	3	24	1	43.6	
8	Divorced /	5	1.5	3	2	1	3.6	
9	Common L	9	2.6	8	1	2.8	1.8	
10	Unknown	0	0	0	0	0	0	
11								

THE CBMS NETWORK COORDINATING TEAM

DR. CELIA M. REYES
CBMS Network Leader

ANNE BERNADETTE E. MANDAP
JASMINDA A. QUILITIS
JOEL E. BANCOLITA
JUAN PAULO M. FAJARDO
MARSMATH A. BARIS, JR
ALELLIE B. SOBREVINAS
JEREMY L. DE JESUS
ERICA PAULA S. SIOSON
NOVEE LOR C. LEYSO
Research Staff

Inquiries regarding this CBMS work may be sent to:

CBMS Network Coordinating Team
Angelo King Institute for Economic and Business Studies
10th Floor Angelo King International Center
Estrada Corner Arellano Avenue, Malate, Manila, Philippines 1004
Tel. No.: (632) 5262067; (632)524-8888 loc 274
Fax No.: (632) 5262067
E-mail: cbms@benilde.edu.ph
cbms.network@gmail.com
Website: www.pep-net.org