

DATA ENTRY SUPERVISION

This guide explains the responsibilities of the data entry supervisor for processing a DHS survey. Once questionnaires have been returned from the field and screened by the questionnaire administrator and office coder/editor, the data entry supervisor's work will commence as explained below.

The data processing activities of a DHS survey involve manual and automatic processes that have a direct impact on the quality of the data. For this reason, all of the staff dealing with these processes should have a clear understanding of their duties as well as the procedures and methods to be used in order for the data processing to run smoothly.

Although the DHS data processing system is designed to work with and without computers connected through a network, it is highly recommended to use a network. In either case, one computer is used by the data processing supervisor to centralize the data (often the server when a network is used) with additional computers used for data entry and verification by the data entry staff. The entire DHS data processing system was written using the software CPro. CPro is being developed as a joint effort between the U.S. Census Bureau, ICF Macro and Serpro S.A.

After installing CPro on all the machines, it is necessary to create a directory structure for both the data entry and supervisor machines. The directory structure is normally established from the root of the C:\ drive in the folder C:\kh61. The directory structure for the data entry computers is as follows:

\ENTRY	To keep the data entry applications
\DATA	To store the main data files for each cluster
\VERI	To store the verification data files for each cluster
\DICTS	To keep the dictionaries needed by the data entry applications
\DOCS	To keep documents related to data entry

The directory structure for the supervisor's computer (and/or server) is as follows:

\ENTRY	To keep the data entry applications
\DATA	To collect the main data files coming from the data entry machines
\VERI	To collect the verification data files coming from the data entry machines
\DICTS	To keep all dictionaries needed by the data processing system
\SUPERV	To keep all applications related to data processing supervision
\WORK	To store main data files after a cluster is accepted
\RAWD	To store raw data files after the verification process is complete, but before secondary editing begins
\FINAL	To store final data files after secondary editing is complete
\REPORTS	To store all reports produced by the data processing system
\DOCS	To keep documents and manuals related to the country's data processing operation

There are several CPro applications used to support the data entry operation; the data entry operators use the following programs:

Entry_menu.bch	Menuing system for the operators
khEntry.ent	Program that allows data entry and verification of the questionnaires

Whereas the following programs are used by the data entry supervisor to facilitate their work:

Superv_menu.bch	Menuing system for the supervisor
Centry.ent	Program that allows manipulation of the control file
khCkld.bch	Program that checks the data structure and checks the number of questionnaires entered in the data files
khEdit.bch	Program that generates the secondary editing listing
Comparer.cmp	Utility that finds keying differences between the main and verification files
khFldCk.bch	Program that produces the check field tables
StatRep.bch	Program that produces the status report on data processing activities

There are five different types of data processing activities that require personnel to execute them. The number of individuals needed to complete each task is closely related to the sample size and the duration of fieldwork. The five types of individuals required are:

- [1] Data processing supervisor: Manages the entire data processing operation
- [2] Questionnaire administrator: Receives questionnaires from the field and archives clusters once they are fully processed
- [3] Office editor/coder: Reviews questionnaires and codes open-ended questions
- [4] Data entry operators: Enters and verifies the questionnaires
- [5] Secondary editor: Reviews the secondary editing reports

The data processing operation can be compared to an assembly line where the unit of work is a cluster. This means that once a cluster arrives in the central office the cluster has to be kept as a whole. One office editor will be responsible for editing and coding a complete cluster; one data entry operator will enter all of the questionnaires for that cluster, a different data entry operator will reenter all the questionnaires for the cluster during verification, etc. Following is a list of the steps involved in data processing:

- [1] Reception and verification of questionnaires
- [2] Office editing and coding of open-ended questions
- [3] Data entry and verification
- [4] Secondary editing

I. RECEPTION OF QUESTIONNAIRES

When a cluster of questionnaires arrives in the office, it will be checked by the questionnaire administrator. The administrator will ensure that all questionnaires and forms have been returned from the field by comparing them to the enumerator assignment sheets. No incomplete clusters should be accepted, and the arrival of incomplete clusters should be reported to the survey manager who may decide to send the cluster back to the field for completion. When an occasional incomplete individual questionnaire is missing, then the administrator should create one by filling out a household or individual cover sheet, assigning a 9 or 7 'questionnaire missing' respectively for the result code. Again, the survey manager should be informed if questionnaires are frequently missing.

Finally, the questionnaire administrator will record the number of completed and incomplete questionnaires on the Cluster Control Sheet for each cluster (see Appendix A for an example of this form) and will then give the cluster with all accompanying forms to the office editor for coding and review.

For more information regarding the role of the questionnaire administrator, refer to the document **khOffice.doc** in directory \DOCS in the supervisor machine.

II. OFFICE EDITING AND CODING

Because data entry will be carried out using an intelligent data entry program, detailed editing is not recommended. Instead, the office editors should concentrate on the checks outlined in the **Office Editing Guidelines** and on the coding of open-ended questions. The office editor must also ensure that the households within the cluster are stacked in ascending order by household number, and that the individual questionnaires within each household are stacked in ascending order by person's line number in the household roster.

Complete editing guidelines are provided in a separate document for the secondary editor. The **General Editing Rules** are as follows:

- The coding of open-ended questions or modifications to the questionnaire made by the office editor, supervisor, and data entry staff should be recorded using a orange and purple pen. The ink color used by the office staff must be different from the color used by the interviewers (blue), field supervisors (red) and field editors (green), so that the data source is always known.
- When making corrections, **never** write over the response initially recorded in the questionnaire. Instead, cross it out with a single line and write the new value alongside the original response.
- As a general rule, when a question requires a value and a response was not written for that question, a MISSING code should be assigned to that question. However, if there is enough evidence to support the assignment of a valid code, then the MISSING value should not be used and the valid code should be assigned instead. MISSING codes are 9, 99, 999 and 9999, depending on the number of digits of the question. If the question allows alphabetic codes, the missing code is the question mark '?' character.

For more information regarding office editing before data entry, refer to the document **khOffice.doc** in directory \DOCS in the supervisor machine.

III. DATA ENTRY AND VERIFICATION

After questionnaires have been reviewed in the office by the Questionnaire Administrator and Office Editor/Coder, they are ready for data processing. However, before this can begin, the batch of questionnaires for the cluster should be registered by the data processing supervisor in the data processing CONTROL file and the cluster should be assigned for main data entry to a data entry operator. To access the control file, use option **C** ("Update CONTROL file") of the supervisor's menu (as described in Section **V** below). Do not register or assign a cluster to a data entry operator until the questionnaires have been reviewed by the questionnaire administrator.

Once a cluster is assigned for main entry in the control file, the batch of questionnaires is passed to the data entry operator. The data entry operator will use option **A** ("Add data to the data file") of the data entry menu to perform this task. The data is stored in the \DATA directory on each data entry machine. It is the responsibility of that data entry operator to enter all questionnaires for that cluster. Once data entry for a cluster is started on a particular computer, it must be completed in that same computer. After entering all questionnaires for the cluster, the data entry operator will return the questionnaires to the data processing supervisor. Upon exiting the data entry program and returning to the data entry operator's menu, the data file is transferred automatically from the \DATA folder on the data entry computer to the \DATA folder on the supervisor's machine.

The supervisor then uses option **A** ("Accept cluster"), invoking the khCkld.bch program, to accept the cluster. This option checks the structure of the data and verifies that all complete and incomplete questionnaires were entered. This is done by comparing the number of questionnaires registered in the control file with those actually present in the cluster data file. If no discrepancies are found, the data file is copied from the \DATA to the \WORK directory in the supervisor's machine and the cluster can be assigned for verification. The supervisor will use option **C** of the supervisor menu to assign a data entry operator to verify the cluster.

This data entry operator will use option **R** ("Re-enter data for verification") of the data entry menu to verify the cluster. It is DHS policy to verify 100 percent of the data. The verified cluster data is stored in the \VERI directory of the data entry machine. The verification must be done by a different data entry operator and on a different computer than the one used for main data entry. After the data entry operator completes the verification, the cluster is returned to the data processing supervisor. Once again, upon exiting the data entry application and returning to the data entry operator's menu, the verified data is transferred automatically from the \VERI folder on the data entry computer to the \VERI folder on the supervisor's machine.

The supervisor then uses option **V** ("Verify double entered questionnaires") of the supervisor's menu to invoke the Comparer.cmp program that compares the main to the verified data files. The result is a report showing the differences in keying between the two files. The file will be loaded into TextViewer and from there the supervisor can print the report.

The report will have three columns; the first column lists the question number, according to the unique name given for it in the dictionary file. The second column contains the response keyed by the first operator, and the third column contains the verifying data entry response to this question. Either the supervisor, a person specifically assigned to this task or a disinterested third keyer, must refer to the original questionnaire to determine which response is correct. When s/he finds a mistake, the original value should be struck out with a single line and the correct response should be written to the right of it.

The cluster along with the report is returned to the data entry operators to fix the discrepancies. Using the data entry menu, the main data entry operator will use option **M** ("Modify the data file") to correct the errors marked in the second column. The verification data entry operator will use option **U** ("Update verification data") to correct the errors marked in the third column. After both files have been updated, the supervisor will again use option **A** ("Accept cluster") to (re)accept the main cluster data and then use option **V** to compare the files once more. This process is repeated until all differences are gone. When there are no differences between the two files, the program will automatically update the control file entry for the cluster to reflect that the verification has been completed.

The original raw data is the data file just after it has been verified and checked for completeness, but before it has been edited. Saving the data at this time, which is a copy of the data exactly as it was collected in the field, allows analysis of the types of field errors encountered. To make a backup of the raw data, the supervisor will select option **K** ("Keep original raw data before editing") from the supervisor menu. The main data file and OTHERS answers file will be copied to the \RAWD directory; however, the index files will not be copied over. The program will automatically enter the date for the "Raw Data Backup" in the control file for the cluster. The supervisor should also mark this on the Cluster Control Sheet.

For more information regarding the data entry operation, refer to the document **khEntry.doc** provided in the data entry menu and also in directory \DOCS in the supervisor machine.

IV. SECONDARY EDITING

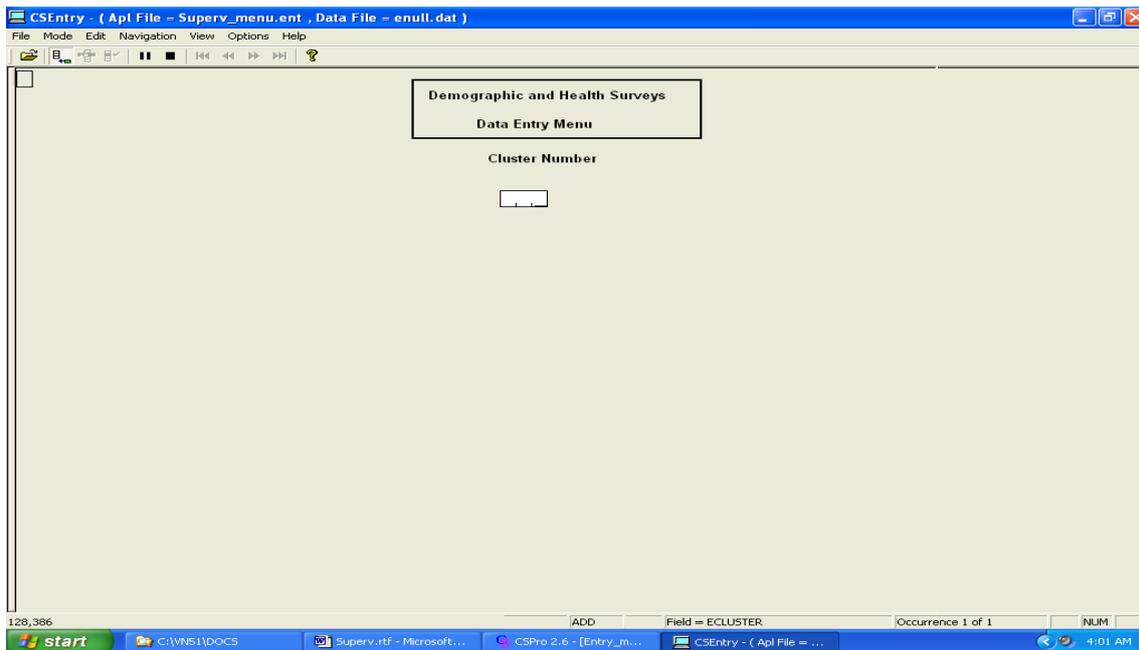
After the raw data is backed up with option **K**, the supervisor will use option **G** (“Generate error listing (secondary editing)”) to run the secondary editing application (application khEdit). The data entry application can be designed in such a way that it will be able to identify any inconsistency at the time of entering the data. However, there are some inconsistencies that require a more in-depth analysis, as well as the attention of subject-matter specialists or senior staff to properly resolve them. Therefore, DHS chooses to make these checks after data entry is complete in a secondary editing application. The result of running this application is an output reporting the inconsistencies identified. The report should be printed and passed along with the batch of questionnaires to the office editors or secondary editing personnel. The office editors will analyze the messages and decide whether data needs to be changed or not. The secondary editor will use the “**Secondary Editing Guidelines**” manual to properly resolve the inconsistencies identified.

If changes are required, the supervisor will use option **M** (“Modify data (secondary editing)”) of the supervisor menu to make the changes requested by the secondary editor. These changes are made using the same data entry application used by the data entry operators and the cluster file residing in the \WORK directory. After making the corrections, the supervisor will use again option **G** to run the secondary editing application. If there are no more error messages generated or the ones remaining were marked 'OK' on the earlier listing, the cluster can be closed. If on the contrary there are still messages outstanding the output along with the cluster package should be passed once more to the office editors. The dates of up to four iterations for secondary editing are automatically recorded in the control file.

If the secondary editor decides that no more changes are required or if after running the secondary editing application no more messages are displayed, the data processing supervisor will use option **S** (“Save final raw data after editing”) to save the final version of the data file. With this option the cycle for the cluster is completed, and the cluster package should be properly archived by the questionnaire administrator. As a result of executing this option, the cluster file is copied from the \WORK folder to the \FINAL folder on the supervisor's machine, and the date for the “Final Data Backup” is automatically assigned to the control file. The supervisor should also write the date on the Cluster Control Sheet.

V. SUPERVISOR'S MENU

The supervisor's menu is used to perform the options described in the previous sections and other options not directly related to a particular cluster such as the production of field-check tables and the data processing status report. To access the supervisor's menu, double-click the DHS Supervisor icon on the desktop. A screen requesting a cluster number like the one shown below appears.



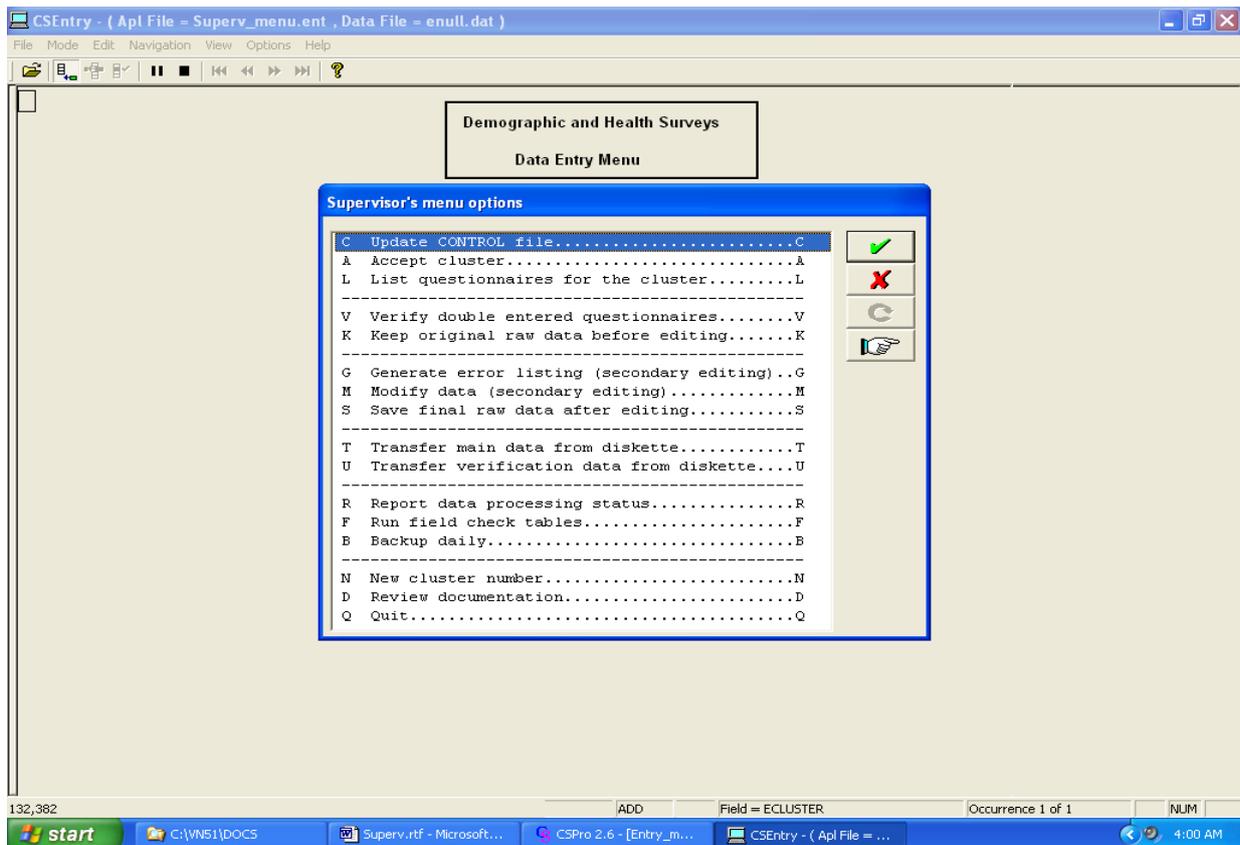
After entering a valid cluster number, the supervisor’s menu (shown below) will appear. The supervisor menu uses the CONTROL file to validate that the cluster given is part of the sample design for the survey. There are several ways to select an option in that menu:

- Type the letter defining the option
- Use the up and down arrows to place the solid bar on top of the desired option and then press enter
- Use the mouse to click the desired option

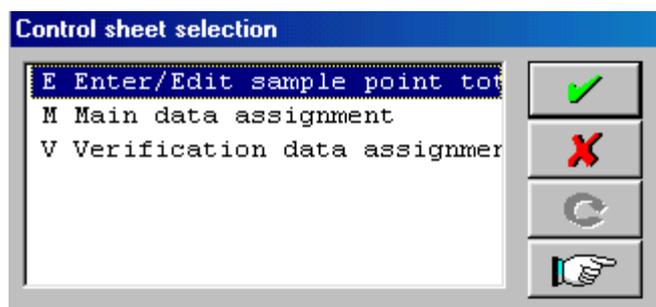
If an option is not acceptable a message will be displayed. The message has to be cleared before proceeding. To clear a message press the function key **F8**. An option is not valid if it is not executed in the proper order. For example, option **K** (“Keep original raw data before editing”) cannot be used until the cluster have gone through double entry and all differences between the main and verified data files have been resolved. The sequence of steps for data processing operations is automatically maintained by the system with the help of the control file.

All reports resulting of the selection of options in the supervisor menu are displayed using the CSPro program “Text Viewer”. This program allows printing reports using the File-Print menu or the program printer icon. Before proceeding with any other options it is necessary to quit the “Text Viewer” program. To do that use the File-Exit menu or click on the exit icon **X** of the program.

After a general description of the data processing operation, each option of the supervisor menu will be discussed:



C Update CONTROL file. This option is used to access the control file in order to enter the totals of complete and incomplete questionnaires and to assign data entry operators for main data entry and verification. After selecting option **C**, the control file menu is displayed.



Option **E** ("Enter/Edit sample point totals") of the control file menu, is used to enter the total number of questionnaires in the cluster. This option also allows the assignment for main data entry.

Option **M** ("Main data assignment") is used to assign the cluster for main data entry. The name and code of the operator, as well as the date of assignment should be provided when using this option.

Option **V** ("Verification data assignment") is used to assign the cluster for verification. The name and code of the operator, as well as the date of verification assignment, should be provided when using this option. Remind the data entry operator to use option **R** in the data entry menu to properly verify the data. This will ensure that the verification

data do not get mixed up with the main data. During entry of verification data, the "OTHER" answers will NOT be requested by the data entry application.

While entering dates in the control file, the application will assign the current date, when the enter key is pressed without providing any other information.

After using any of the options from the control file menu, the program will request to confirm to update the data. Press **Y** (Yes) to update, or **N** (No) to quit without saving the changes, or **R** (return and correct) to make a change to the data just entered. Once any of these options is chosen the application will return back to a screen displaying a cluster number. Exit the screen by using the File-Exit menu or by clicking on the closing screen icon **X**.

- A Accept cluster.** This option should be used after main data entry for a cluster is completed. Once a data entry operator completes a cluster, the data is automatically transferred to the computer used by the data-processing supervisor or a server.

The cluster will not be accepted if the number of complete and incomplete questionnaires does not match the count registered in the control file. The option will produce a list of questionnaires as well as totals from both the actual cluster data file and the control file. If the counts are the same, the data is ready for verification and the system will automatically update the control file. If the counts are not the same, print the listing of questionnaires. Use this listing to check the data entered with the questionnaires. If some questionnaires were not entered, return the cluster to the data entry operator so that missing questionnaires are added. However, if the mistake occurred in the counting of the questionnaires, correct the control file using option **C** as explained above. Make sure to update the Cluster Control Sheet as well.

Once a cluster has passed the verification check, the supervisor's computer will no longer accept that cluster from a data entry operator's computer.

Program: khCkID
A list of household and individual questionnaire id's with result codes and total count of questionnaire as well as error-messages are displayed by "Text Viewer".
Actions: If no errors are found data is copied from directory \DATA to directory \WORK.
Control file variable CMCHECK is updated to reflect acceptance of the cluster.

- L List questionnaires in data file.** This option will list all of the household and individual questionnaires for the cluster. The option uses the same application that is used by the previous option.

Program: khCkID
A list of questionnaires along with totals is displayed by Text Viewer.

- V Verify double entered questionnaires.** This option will allow the comparison of data entered in the main data file and in the verification data file.

Program: COMPARER
A list of discrepancies between the main and the verification files is displayed by Text Viewer.
Actions: If no discrepancies are found, control file variable CVERIF is updated to reflect that the cluster passed the verification stage.

K Keep original raw data before editing. This option allows a backup of the cluster data after the cluster passed verification, but before any secondary editing corrections have been made to the data.

Program: khCkID
A list of questionnaires along with totals is displayed by Text Viewer.
Actions: If no errors are found data is copied from directory \WORK to directory \RAWD.
Control file variable CRAWDATA is updated to reflect that a raw data backup has already been done.

G Generate error listing. This option will generate the error listing for the secondary editing.

Program: khEdit
A list of inconsistencies is displayed by Text Viewer.
Actions: Control file variable CEDIT is updated to reflect that cluster already passed the secondary editing stage.

M Modify data file. This option is used for correcting errors found in the data file as part of the secondary editing process. To aid the supervisor and data entry operators regarding variable names and labels when trying to resolve problems on the questionnaire during data entry or correction to the data, please refer to **Appendix B**.

Program: khEntry
Actions: Cluster data file in directory \WORK will be modified.

S Save final edited data file. This option will save the final edited data file on the hard disk. After all corrections have been made to the data, and the data are clean, the final data must be saved to ensure no loss of data. This is the end of data processing for this cluster. Store the questionnaires for the cluster in the storage room, with the clusters ordered according to cluster number, so that they can be retrieved easily in the future.

Program: khCkID
A list of questionnaires along with totals is displayed by Text Viewer.
Actions: If no errors are found data is copied from directory \WORK to directory \FINAL.
Control file variable CFINDATE is updated to reflect that the cluster has completed the data processing cycle.

T Transfer main data from diskette. In case a network does not exist or a machine is not connected to the network, the transfer will be done by diskette from the data entry machine to the supervisor's machine.

U Transfer verification data from diskette. Same as option T except that verification data are being transferred from data entry machine to the supervisor's machine via diskette.

R Report data processing status report. This option will generate a report which shows the progress in the processing of each cluster and of the survey as a whole. As this report is not part of any particular cluster, it can be produced any time regardless of the cluster being selected when accessing the supervisor menu.

Program: STATREP

The results will be shown on the screen by the TextViewer program and they can be printed with the File-Print menu or the printer icon.

- F Run check field tables.** This option will generate a list of tables from all the data entered up to the current date. Field check tabulations should be produced on a weekly basis, at the end of the week. Copies of them should be given to the survey director and sent by E-mail to Macro. These tables allow to review the progress of fieldwork and to help identifying problems with interviewers and teams. Corrective actions can then be taken. The option concatenates single cluster data files in the supervisor's computer into a single data file, and will produce a set of tabulations. The tables should be converted to an RTF file using the "Tables Editor" software. Once the tables are in RTF format they can be printed using "word". As with the status report, this option can be used at any time regardless of the cluster being selected when accessing the supervisor menu.

Program: khFldCk
A TBD file will be produced to be accessed later by the "Tables Editor" program.

Actions: Clusters files will be concatenated from directory \DATA\C???.DAT into file \REPORTS\KHIQ60.DAT
The TBD file should be uploaded by the table's editor and using the Options-Save All to RTF menu converts the tables to RTF.
Load the RTF file into word and print it.

- N New cluster number.** This option allows changing the current cluster being processed to a different one.
- B Backup daily.** This option is used to back up the data files to another hard disk or to an external storage device on a daily basis.
- D Review documentation.** This option may be used to display this document. The document is called **khSuperv.doc** and it is located in directory \DOCS in the supervisor machine.
- Q Quit.** This will quit completely from the supervisor's menu. Always use this option to leave this menu; it is designed to prevent concurrent sessions of the supervisor menu.

Appendix A

Cluster Control Sheet for Cambodia 2010

Cluster <# # #>

	Domain		<name>
	Province		<name>
	Urban/Rural		<name>
	Total	Complete	Incomplete
Households	_____	_____	_____
Women	_____	_____	_____
Men	_____	_____	_____
	Main entry	No.	Verification
Operator	_____	_____	_____
Computer	_____		_____
Date	_____		_____
Checked	_____	Verified	_____
Raw backup	_____		
Editing	_____		

Final backup	_____		

○
Main

○
Verification

○
Verified

○
Edited

Cluster <# # #>

Appendix B

CSPro Print Preview of dictionary variable names and labels.