

DATA ENTRY GUIDELINES

1. GENERAL GUIDELINES FOR DATA ENTRY

The data entry for the DHS survey will be carried out using the software package CPro. CPro is an interactive data entry system that can check for acceptable codes for questions, follow skips and filters in the questionnaire and check the consistency of data as they are entered. The DHS questionnaires will be entered by cluster, with each cluster being assigned to one data entry operator. The data for each cluster will be entered into a separate data file for that cluster on the hard disk of the computer, rather than into one large data file. This is to protect against a major loss of data due to hardware or software failure.

During data entry it is important to remember that speed of data entry is not important but accuracy of entry is important. CPro is like a computer-aided office editor. It will not allow any values to be entered that it thinks are out of range. It will skip to the next question that it believes is the next to be entered. It will present a new screen when it thinks a new screen is necessary. CPro does these things according to the data entered. It can not check everything that the operator enters. If an operator enters a value that was not recorded on the questionnaire, but is a value that is within range, CPro will go to the next field that this value demands. Each operator must review what they have entered as they are entering data to ensure that mistakes do not take place.

Because CPro will control which question is the next to be entered, it is important for the operator to watch the screen, to see which field CPro expects to be entered next. Operators should not enter data with their heads down, only reading the questionnaire. This will lead to many errors in the data entered if the operator miskeys any data or if there are any mistakes in the responses recorded on the questionnaire. At a minimum, the operator should look at the screen every time a page is turned in the questionnaire.

Data are recorded on the questionnaire in two ways: (1) codes are circled, or (2) numbers are written in boxes. Some questions allow the response to be in either of the two forms or in a combination of the two forms. For example, «A104» uses circled codes, «A101» uses codes written in boxes, «A105» uses either codes written in boxes or circled codes, and «A237» uses a circled code followed by a code written in boxes or a 3 –digit circled code.

If during data entry the operator finds that there are data in the questionnaire which she/he has not been asked by CPro to enter or she/he finds that CPro is demanding an answer to a question that was not asked in the survey, then the operator should review all of the data entered to ensure that she/he has not entered the wrong value in one of the questions. When reviewing the data entered check question by question whether the data entered is the same as the response written on the questionnaire. Do not look at just the last few questions, but look back over the whole page and the preceding page to ensure that no typing error was made in an earlier question.

If there was no mistake by the operator in entering data then she/he should ask the data entry supervisor to resolve the mistake. If the supervisor is unable to resolve the mistake then the problem should be shown to the senior survey staff for resolution.

During data entry, messages may be produced which will require attention and correction to the data entered. Operators should read the message carefully before attempting to resolve the problem. Resolve the problem in the same way as before, reviewing the data entered first to check that no entry error was made. If no entry error occurred then ask the supervisor to resolve the problem. The supervisor will have a manual containing a list of messages and ways to resolve the problems.

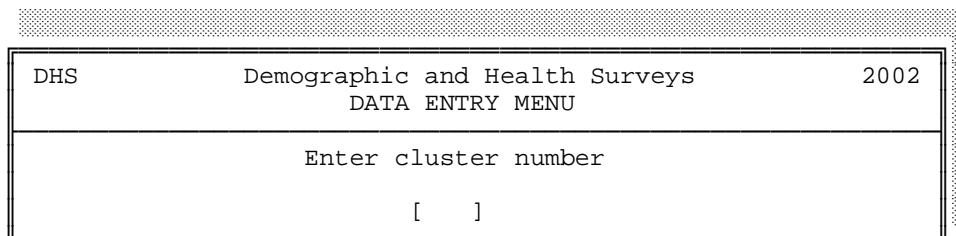
During data entry keep all questionnaires in a tidy pile. Once a questionnaire has been entered turn the questionnaire over onto a second pile. The questionnaires should be kept in numeric order by household at all times, and within households individual questionnaires should be kept in order according to line number of the respondent. At the end of data entry for a cluster re-tie the pile of questionnaires with the cluster control sheet on top and pass the cluster to the supervisor, who will then assign a new cluster for entry.

2. STARTING DATA ENTRY

a) Data Entry Menus

To start CSPro data entry menu:

1. Make sure your computer is on and the Windows desktop is displayed.
2. Click the DHS Data Entry icon on the Windows desktop.
3. The following screen will appear, requesting the cluster number be entered:



DHS	Demographic and Health Surveys	2002
DATA ENTRY MENU		
Enter cluster number		
[]		

4. After typing in the cluster number, the Data Entry menu screen below will appear:

DHS Demographic and Health Surveys	2002	DATA ENTRY MENU	Option code
Available options			
Add data to the data file.....			A
Modify the data file.....			M
Transfer data to diskette.....			T
Re-enter data for verification.....			R
Update verification data.....			U
Transfer verification data to diskette.....			V
New cluster number.....			N
Review documentation.....			D
* Enter an option code (Esc to quit)			[]

The menu shown above gives various options:

A Add data to the data file. This option is used to enter data for the first time. On the cluster control sheet, this is called the “Main” Data entry.

M Modify the data file. This option allows corrections to be made to questionnaires already entered into the “Main” data file.

T Transfer data to diskette. This option allows the transfer of main data from the data entry machine to be transferred to a diskette once a cluster has been entered.

R Re-enter data for verification. This option is used to enter data that are being verified or entered for the second time. It is identical to option A, except that a different data file is being used for the questionnaires.

U Update verification data. This option is used to modify the verification data entered with option R. It is identical to option M, except that it uses the verification data file.

V Transfer verification data to diskette. This option allows the transfer of verification data from the data entry machine to be transferred to a diskette once a cluster has been entered for verification.

N New cluster number. This option will be used to change or correct the cluster number that is currently being processed. When data entry for a cluster has been completed, this option is used to change to a new cluster number.

D Review documentation. This option will display this documentation on the screen and allow the user to browse through the documentation.

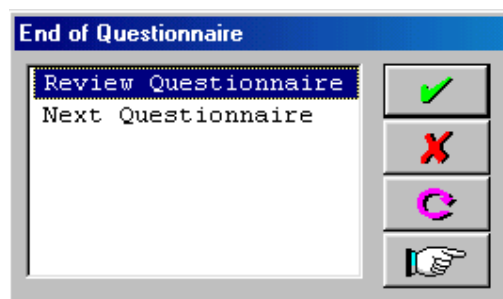
If you are entering data into the “Main” data file, press option A; if you are verifying data, press option R. The first data entry screen will be displayed.

Press the Num Lock key by the numeric keypad to the right of the keyboard so that the Num Lock light is shining (if a light exists). This will permit the use of the numeric keypad for the entry of data. The numeric keypad should be used for all numeric data entry. The alphabetic keys will only be used to enter textual responses specified against the category "Other" and for the questions with alphabetic codes. During data entry, enter the values shown on the questionnaire including all leading zeros.

b) Household Schedule Cover Sheet

Each screen is laid out in a similar manner to the questionnaire. On the first screen enter the «AHCLUST»Cluster Number, «AHNUMBER»Household Number, «AHREGION»Region, «AHTYPE»Urban/Rural, and «AHLOCAL»Large City/Town/Village codes in order. The program will automatically go from one field to the next field when a field is full; it is not necessary to press ENTER after entering each field. After the Cluster Number and Household Number codes have been entered the program will check to see if this household has already been entered. After this, continue with the «AHINTD»Day, «AHINTM»Month and «AHINTY»Year of Interview, «AHINTNUM»Interviewer Number, and «AHRESULT»Result Code. If the Result Code is 1 (Completed) then enter the «AHVISITS»Total Number of Visits, «AHMEMBER»Total in Household, «AHWOMEN»Total Eligible Women, «AHMEN»Total Eligible Men (if applicable), and «AHRESP»Line Number of the Household Respondent. If the Result Code is not 1 then the program will skip these fields. Finally enter the «AHSUPERV»Supervisor code, «AHFEDIT»Field Editor code, «AHOEDIT»Office Editor code and the «AHKEYER»Keyer code. The Keyer code is the operator code assigned to each operator. If you are entering data into the “Main” data file, your keyer code should be written on each questionnaire. If you are verifying data, only enter your keyer code but it should not be written on the questionnaire.

If the result code was not 1 then that is all that will be entered for that questionnaire and CSPro will display the following:



First check that there are no more data to enter for the questionnaire before continuing. By typing the letter R (Review Questionnaire) the data entered can be

reviewed and corrections made, if necessary. If the letter N (Next Questionnaire) is selected, CSPro will then show a new household cover sheet for the next household questionnaire. Only continue with the next questionnaire after checking each of the fields to ensure that they are correct.

If the result code was 1 then CSPro will show the next screen for the household members and give the line numbers of each of the household members that it is expecting to be entered.

c) Household Schedule

Check that the number of lines with line numbers on the screen is the same as the number of lines completed on the questionnaire.

Enter each member in order, entering data across the lines. When the eligibility columns (columns 8, 9) are reached enter the number from that column if it has been circled, but otherwise enter 00 if it has not been circled (CSPro will automatically put 00 in for persons who are not eligible in most cases). If there is a male sample, there will be another column for eligibility. Continue entering the members until all members have been entered. After entering the last data for the last member, CSPro will present a similar screen to the one shown above. Check all of the entries on the page to ensure that they have been entered correctly, and then select the letter R to review or correct data on the schedule, or the letter N to go to the next section.

After completing the next section, relating to the characteristics of the household, CSPro will present the screen for the height and weight of women. Check that the number of lines for which numbers are showing is the same as the number of lines completed on the questionnaire. If there are no eligible women in the household, this section will not be displayed.

After completing the height and weight section for women go on to the height and weight section for children. Check that the number of lines for which numbers are showing is the same as the number of lines completed on the questionnaire. If there are no eligible children under the age of 6 in the household, this section will not be displayed. In this section, questions 36-39 are for all children under the age of 6, however questions 40-43 are only for children born in «D1997» or later.

After the height and weight sections, we are ready to enter the first individual questionnaire if there are eligible respondents.

If there are no eligible women and no eligible men in the household then CSPro will present the screen for the next household questionnaire, but if there are eligible women or eligible men, CSPro will display the screen for the Individual Questionnaire cover sheet. If the Household Schedule cover sheet is shown instead of the Individual Questionnaire cover sheet and there are individual questionnaires to be entered, ask the supervisor for assistance in correcting the problem.

d) Individual Questionnaire Cover Sheet

When the Individual Questionnaire cover sheet is displayed the «ACLUSTER» Cluster Number, «ANUMBER» Household Number, «AREGION» Region, «ATYPE» Urban/Rural, and «ALOCAL» Large City/Town/Village codes will already be displayed. Check that each of these is correct before continuing, and then enter the «ALINE» Line Number of the Woman or Man. The individual questionnaires must be entered in order according to that line number. Next enter the remaining fields on the Individual Questionnaire Cover Sheet.

If the result code was not 1 then that is all that will be entered for that individual questionnaire and CSPro will again display a screen, asking for the data to be reviewed or to continue with the next questionnaire. Only continue after checking that each entry on the page is correctly entered.

e) Individual Questionnaire

Continue entering data in the same manner for each of the questions in the questionnaire, verifying that the correct route is being followed at all times and that no fields are being skipped that have data in the questionnaire. Throughout the entry of the questionnaire, ensure that the data on the questionnaire is the same as the data entered. Do not try to enter the data too quickly. It is important that the data are correct and speed of data entry is not crucial.

There are certain key sections of the questionnaire that will require special attention and these are explained in further detail in section 8.

At the very end of the questionnaire, CSPro will again display a screen allowing the data entered to be reviewed or to continue with the next questionnaire. Only continue after checking that the data entered are correct.

After an individual questionnaire is complete, CSPro will display the cover sheet screen for an individual questionnaire if there are more individuals to enter for the household, or the cover sheet for the household if the individual just entered was the last individual in the household.

3. MOVING BACK THROUGH THE DATA

If a mistake is made during data entry then it is possible to go back and correct the mistake provided the mistake is found before the end of the questionnaire. It is possible to go back to questions on previous screens as well as to questions on the current screen. There are two ways of going back to earlier questions:

- a) Pressing up arrow or Shift Tab key to go back one field. Repeating this will go back another field. This can be used to move back to the field where the correction is to be made.
- b) Pressing the function key F6 (GoTo) will ask for the name of a field to go back to. Enter the name of the variable to correct and then click OK to go to that field. If the field is part of a repeating section of the questionnaire then

an occurrence number can be given by pressing the Tab key after giving the name of the field, entering the occurrence number and then clicking OK.

In the individual questionnaire the variable names are based on the question numbers, preceded by the letter Q and sometimes followed by another letter if there is more than one field for the question. Generally the letter after the question number will follow the pattern A, B, C, D... although there are some exceptions. If the question asks for months and years then the two parts are suffixed M and Y. If the question asks for hours and minutes then the two parts are suffixed H and M. If the question asks for units and numbers then the two parts are suffixed U and N. Some examples are «A101H», «A101M», «A105M», «A105Y», «A628A», «A628B», «A628C», «A628D», «A237U», and «A237N». A list of all variable names is available with the data entry supervisor. Names of variables in the household questionnaire cover sheet and on the individual questionnaire cover sheet are given by name and the list of variables should be used to refer to these variables. Variables in the household schedule and the other sections of the household questionnaire consist of QH followed by the number of the question, e.g. «AH03» or «AH16». Variables in the men's questionnaire consist of QM followed by the number of the question, e.g. «AM103».

4. CORRECTING DATA

If a mistake has been made in the middle of entering data into a numeric field, pressing the Space bar can erase the information already entered in that field. For example, if code 01 was to be entered for «A307» but the 1 was pressed first then the whole field can be erased by pressing the SPACE bar. The corrected data can then be entered into the field. The Backspace and Del keys cannot be used to correct data entered in numeric fields.

If a previously entered numeric field contains a mistake, the correct data can be entered over the top of the previously entered data. However, if the operator is not sure, she/he should use the SPACE bar to empty the field before entering the data.

In correcting data in alphabetic fields, the Backspace, Del and Insert keys can be used. The Space bar may be used to enter a single blank character in these fields, but does not completely blank out the field.

5. MOVING FORWARD THROUGH THE DATA

Having moved back through the data to an earlier field there are two ways of moving forward again.

- a) Press the down arrow key or Tab key for each field until the required field has been reached.
- b) Pressing the function key F6 (GoTo) will request the name of a variable to advance to. Giving the name of the variable and then clicking OK, as for moving back through the data, will move to the variable requested. If it is desired to move to the last variable entered in the questionnaire, press F10

(Advance to end) without specifying the name of a variable. It is not possible to move forward further than the last variable entered, so CSPro will advance only as far as this variable.

6. SKIP ERRORS AND MISSING DATA

CSPro is programmed to present on the screen each of the questions that should be keyed by the operator. As the response for each question is entered, the operator creates a data path through the questionnaire. Certain responses are used to help control this path. For example, prior to entering the household schedule, the data entry operator will be asked to enter the total number of persons in the household. This field is then used to control the entry of the household schedule. Similarly, the total number of eligible women (and/or men if included) is used to control the number of individual questionnaires entered for the current household.

In the individual questionnaire there are also responses to questions that control the data path. For example the year of birth of each child in the birth history is used to control whether the child should be included in the health section of the questionnaire. Any field that determines a skip is also controlling the data path. For example, «A321» contains several skips: codes 02, 12, 13 and 96 skip to «A331», codes 08-11 skip to «A325». Code 07 skips to «A328».

If the fields that control the data path are miskeyed, the data entry operator may be asked to enter responses that do not exist on the questionnaire or may not be allowed to enter responses that do exist in the questionnaire. This is because the data entry operator is following an incorrect data path. If this occurs, the operator should return to the "control" field that determined the skip and correct the entry for that field.

Occasionally a skip error will exist in a questionnaire. If a woman has reported that she is currently pregnant according to question «A226», but the following question records no answer as to the duration of her pregnancy, the interviewer has committed a skip error. However, an operator should never assume that the previous keying has all been correct and that the error is in the questionnaire. The operator should check that the code recorded in question «A226» has been properly entered.

If the questionnaire has a code 1 (Yes) circled for question «A226», then the error, in fact, is in the questionnaire. In this case the operator, using a GREEN pen, should record a 99 in question «A227»A227 as well as entering 99 in the field on the screen. If, however, question «A226» has a code 2 (No) circled, but the operator has entered a 1, the 1 entered into the screen should be corrected to a 2. Once the 2 is entered, CSPro will no longer demand a response to question «A227».

In general, if a skip error is found in the questionnaire and data are missing the operator records a special code indicating missing data. For numeric fields, this code is a field full of 9s. For alphabetic fields, this is a single question mark "?" recorded in the first position of the field. Any corrections made during data entry should be written into the questionnaire with a green pen.

In another situation, if there is no response recorded on the questionnaire for question «A226», the missing data code, a single 9 in this case, should be recorded on the questionnaire and in the field on the screen. However, this field is used to control the data path, and a missing data code in question «A226» will skip over question «A227». If a response is recorded in «A227», it can be deduced that the only possible response to question «A226» should have been code 1. In this case, the code 1 should be recorded on the questionnaire and on the screen for question «A226».

In general, if a question was left unanswered and the question affects the data path and there is information in a question on one path controlled by that question which indicates that there could only be one possible response to the controlling question, that response should be recorded for the question. For example, if question «A421» was left unanswered, but a response was recorded for question «A422», the code 1 should be recorded in question «A421», as it is the only code that will lead to question «A422».

It is a good habit to review other responses prior to the field that is identified as giving an error. Occasionally a data entry operator can be "off" by one field and enter many erroneous responses without the errors being detected. This can often happen in sections of the questionnaire that contain several Yes/No questions in sequence, e.g. «A412». It is essential for operators to review earlier sections of the questionnaire when an incorrect data path appears to have been followed. If an operator is unsure of the problem, she/he should never hesitate to call a supervisor for help.

Operators should make certain that they enter the responses found in the questionnaire and not create responses in order to satisfy CSPro's demand for a response. A supervisor should always be consulted for those cases where serious skip errors are found in a questionnaire. These errors should also be recorded on the Problem Log kept by each operator and supervisor.

7. DATA PATH

CSPro remembers the current data path at all times and shows all values that are on the current path in green. When moving back through the data CSPro will change the fields after the current field to white. If a correction to an earlier field changes the flow of questions and a field that had data in it is bypassed then this field will become gray. At the end of entry of a questionnaire only those fields that are green are written to the data file, any fields that are not on the correct path are not written to the data file.

The reason for explaining this is that if a mistake is made such that data are entered on the wrong path then it is not necessary to go to those fields and blank out the values entered in them, providing that the final path through the questionnaire is correct.

8. QUESTION DETAILS

a) Alphabetic code fields

There are several fields that have alphabetic codes for the responses and these are either circled or not. Enter all of the codes that are circled for the question and then press the ENTER key. The codes must be entered in order, with no gaps between the codes and no repetitions of the codes. The questions that this applies to are: «A311», «A330», «A407», «A422», «A426», «A443», «A465», «A471», «A474», «A480», «A482», «A496», «A525», «A528», «A607», «A620», «A803», «A818», «A819». If there is a questionnaire for men, check throughout the questionnaire for alphanumeric variables. They should start with the letters QM.

b) Other answers

Each "Other" answer recorded in the questionnaire is to be recorded. When the screen for the other answer is displayed, type in the text of the response exactly as it is given on the questionnaire. Do not abbreviate the text; give as much detail as possible. After entering the text of the response, press enter to return to the data entry screen.

c) Decimal places

Data are entered with the decimal places in all cases. The questions this relates to are: «AW40», «AW41», «AC40», «AC41». Enter each of these fields with great care, always entering all leading zeros, as well as the decimal place. Always check the value entered immediately after entering the data. It is very easy to make mistakes on these questions. Particularly, beware of cases where the value after the decimal point is not completed; we assume that this value should be zero, and so it is necessary to enter the correct number of zeros to fill out the field.

d) Multi-part questions with single codes

Question «A615» has three parts (boys, girls and either), but also allow a single code 96 to be used. If code 96 is entered in the first of the boxes on the screen then CSPro will automatically put the same value in the second and third of the boxes on the screen.

9. SECTION DETAILS

a) Section 2 - Birth History

Before entering data in the birth history, check that the number of lines with line numbers agrees with the number of lines containing data on the questionnaire. Check carefully what order the children are to be entered in as they are often listed out of order, and the line numbers corrected on the questionnaire. The children must be entered in increasing order of date of birth from the oldest child to the youngest child.

When entering data it should be entered across the lines of the birth history, which is child by child.

b) Section 3 – Contraception

Enter the data on this table with care, since interviewer errors occur more often here, circling a response for «A302» when «A301» is not 1 or not circling a response in «A302» when «A301» is 1. Establish a pattern of data entry remembering that a code of 1 in the first column requires a response in the second while a code of 2 in the first column requires no response in the second.

c) Section 4A - Pregnancy and Breastfeeding

Before entering data in this section check that the number of columns completed in the questionnaire agrees with the number shown on the screen. The children should be listed in the questionnaire in order from the youngest to the oldest. If there are more than two column numbers shown on the screen, ensure that there is a continuation questionnaire containing the remaining children's data.

Enter the data column by column, turning to the following page for each child before going back to the start of the section for the next child.

d) Section 4B - Immunization and Health

Repeat the same checks as for section 4A when starting the section relating to immunizations and health and complete the screens in the same manner.

10. CALENDAR

The entry of the calendar is one of the most important parts of the data entry process. It is essential that great care be taken in entering these data. The calendar is constructed of 1 or 4 columns, with a row for each month since January «D1997». The rows are numbered 01 to 72. In the data entry program, the calendar will appear just after the birth history. The columns of the calendar are represented as lines across the screen, with each row marked down the lines. The data are entered column by column, by specifying the first and last rows in which a particular code exists, and then the code that is in those rows. The program will then place the code within the calendar in the correct rows for the column.

Initially the Column field will have the value 1 and the Minimum field will have a value between 1 and 12, representing the row of the calendar for the month of interview. The cursor will be in the Maximum field. Look at column 1 of the calendar and look down the column until the first code changes. In the Maximum field enter the value of the last row that has the same code, and then enter that code in the Code field. The calendar will be filled automatically with the code for the specified rows. Repeat this for each of the codes that appear in the calendar until the end of the column.

If the calendar has more than one column, the data entry program will now change the column number in the Column field to 2, and place the cursor in the Minimum field. Look at column 2 and find the first row in which there is a code in column 2, and enter that code in the Code field. Then find the next code and enter the row number in the Minimum field. The Maximum field will automatically be filled with the Minimum field's row value. Enter the code from column 2 into the Code field. The calendar will be filled automatically. Repeat this until all codes recorded in column 2 have been entered in the calendar.

After completing column 2, hold down the Shift key and press the Tab key to go back to the Column field. Change the value to 3. For each of the codes in column 3, do the same as in column 2.

At the end of column 3, again use Shift and Tab to change the Column value to 4. The Minimum value will be change to a value between 1 and 5, as for column 1. Give the Maximum row value for column 4 for the first code, and then enter the code in the Code field. Repeat this until the end of the column.

At the end of column 4 (or column 1 if only one column exists) the program will go back to the Column field and put a 9 in the field. This value is the Exit code for the calendar. Press Enter to finish the calendar. The program will verify the information entered in the calendar. If the program finds any problems, a message will be displayed, explaining the error, and corrections should be made. Usually the program will set up the Column, Minimum and Maximum fields where it believes the error is to be found. Correct the errors in the calendar, and then use Shift and Tab to go back to the Column field to use the code 9 to exit from the calendar. After all errors have been corrected, the program will continue with the date of termination of the last non-live birth pregnancy prior to January «D1997». After the calendar, continue with the next question in the questionnaire (question «A222»).

11. ENDING QUESTIONNAIRES

When coming to the end of data entry for a questionnaire, review the data entered to ensure that no mistakes have been made. After finishing the entry of a questionnaire there is no chance of correcting the data on that questionnaire until later in a "Modify data" session. If a mistake was made but not discovered until after the end of the questionnaire, record the cluster number, household number, individual's line number, question number (and occurrence number if part of a set of repeating fields), with the error and ask the data entry supervisor to assist you in correcting the problem.

After the end of a questionnaire the data for that questionnaire is not saved until the operator clicks Yes when asked to Accept the questionnaire. This will bring up the first screen for the next questionnaire, i.e. a new household questionnaire screen or a new individual questionnaire screen.

12. STARTING A NEW CLUSTER

When a cluster of questionnaires has been finished, a new cluster will be assigned. Each cluster is to be entered in a different data file. The cluster number to use must be changed. Leave the data entry program by pressing the function key Alt+F4 (File, Exit). This should only be done when the computer is displaying the household questionnaire cover sheet and awaiting entry in the first field of the screen. After ending the entry program, the data entry menu will then be displayed. Use option N to change the cluster number to the new cluster number. Press option A to commence data entry for that cluster if entering data for the first time in the "Main" data file or press option R to reenter data into the "Verification" data file.

13. QUITTING DATA ENTRY

The only point at which it is possible to quit data entry is at the very beginning of a new household schedule cover sheet. Pressing the function key Alt+F4 (Exit CSEntry) at this point will leave the data entry program and return to the data entry menu. Pressing Esc will quit the menu completely and return to Windows desktop.

Never leave the computer in the middle of data entry. If the operator wishes to leave the computer then she/he must finish the household that she/he is entering and then press Alt-F4 to end and return to the data entry menu. **Always leave the computer on the data entry menu.**

As the end of the data entry session or the end of the day approaches, remember the amount of time it takes to enter a household and all of its individual questionnaires. Do not start a new household unless there is time to properly complete the household.