END-DECADE

MULTIPLE-INDICATOR DATA PROCESSING MANUAL

ISSA

00.4.20

Overview

Data processing personnel

The data processing team for a MICS survey includes four types of personnel: questionnaire administrators, data entry operators, secondary editors and a data processing supervisor. Each position has distinct responsibilities and combining them is likely to damage the quality of your data.

Questionnaire administrators

The questionnaire administrators check the clusters that arrive from the field. They check that all of the questionnaires are present and ready to be entered. If there are missing questionnaires, they must contact the fieldwork team and try to find them. If the missing questionnaires cannot be found, the questionnaire administrators must resolve the problem.

In the case of household questionnaires, this means requiring the fieldwork team to redo the household interview. If this is impossible, the questionnaire administrators must add a blank household questionnaire with the result code 5.

In the case of women's and children's questionnaires, this means requiring the fieldwork team to redo the missing questionnaires. If this is impossible, the questionnaire administrators must change the completed questionnaire totals on the household cover sheet.

Data entry operators

The data entry operators enter the data. They should have prior data entry experience and be familiar with the questionnaires. Before beginning data entry in earnest, a two or three day training must be held. By the end of the training, the data entry operators should be comfortable with the data entry program and aware of their responsibilities.

Secondary editors

Secondary editors investigate complex inconsistencies discovered during the data entry process. They must have an excellent grasp of the questionnaires and the goals of the survey. Editing guidelines are provided to aid them during the secondary editing process.

Data processing supervisor

The data entry supervisor oversees the data processing system. The data entry supervisor must have an excellent grasp of the questionnaire and programming skills in the data entry package and SPSS. The data entry supervisor also resolves problems encountered by the data entry operators.

Data processing equipment

The list below shows the equipment that is necessary for data processing.

- Data entry machines
- A supervisor's machine
- A secondary storage device (e.g., a Zip or Jaz drive)
- Diskettes
- A printer
- Paper
- Toner cartridges/printer ribbons

- Surge protectors
- Uninterrupted power supplies
- Green pens
- A data entry room
- An editing room

The data entry machines should have at least 486 processors, Windows 95 or higher, 16 megabytes of RAM or more, 500 megabytes of hard disk space or more and a 3.5'' floppy diskette drive. The supervisor's machine should have a 300 Mhz processor, Windows 95 or higher, 64 megabytes of RAM or more, 500 megabytes of hard disk space or more, a 3.5'' floppy diskette drive and a CD-ROM drive. The secondary storage device should be connected to the supervisor's machine.

The Uninterrupted power supplies are essential if the country in which you are working suffers from power outages. The green pens are for the data entry clerks and the office and secondary editors. They should be used whenever a questionnaire is modified. The green ink distinguishes their changes from the original value recorded by the interviewer and the changes made by the fieldwork team.

The data entry room should be large and cool. Each data entry agent should have space for their machine and the questionnaire they are working on. The editing room is for the office and secondary editors. It must contain shelves or cupboards in which the questionnaires can be stored in an organized fashion.

Data processing design

MICS data processing is organized around clusters. The data for each cluster are entered into a set of data files whose names include the cluster's number. This approach breaks data processing up into discrete segments and keeps the size of data files to a minimum.

A MICS survey has two data processing phases: primary and secondary. The goal of primary data processing is to produce clean, edited data files. Primary data processing comprises the following steps.

- Data entry
- Structure check
- Verification
- Secondary editing
- Backup of data

The goal of secondary data processing is to produce data files to be used for analysis, including tabulations. Secondary data processing comprises the following steps.

- Calculation of weights
- Recoding of variables
- Tabulation

The flow chart on the following page illustrates the flow of primary data processing.
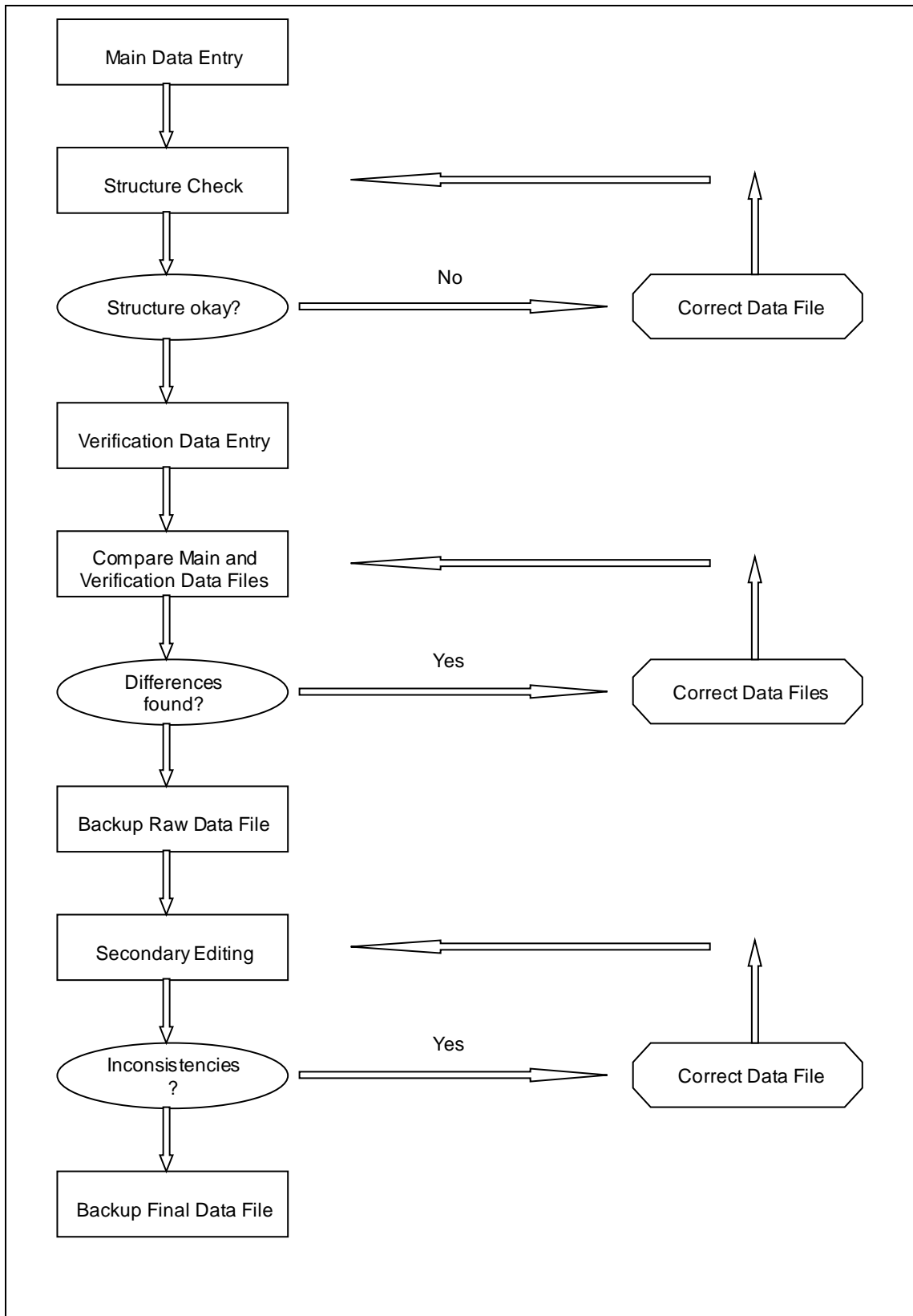
Main data entry

The data for each cluster are entered twice. The first time the data is entered is called main data entry and the second time is called verification data entry. One data entry agent must do the main data entry and another data entry agent must do the verification data entry. Under no circumstances should the same data entry agent enter both the main and verification data.

The data entry program performs basic consistency checks. These checks are designed to catch egregious questionnaire errors and data entry errors. Complex inconsistencies are not addressed until secondary editing.

Structure check

It is essential that all questionnaires be entered and that their identification information be correct. The structure check ensures that both of these requirements are met. Only the structure of the main data entry needs to be checked because any structural errors in the verification data will be discovered during the verification process.

Verification

When main and verification data entry are finished, the two files are compared and a list of differences is produced. The two data entry clerks resolve differences by checking the questionnaires for the correct values. They then correct the values in their data files and the verification program is rerun. The process is continued until the data files are identical. When the verification process is finished, the data should contain exactly what is written on the questionnaires.

Raw data backup

When the data have been verified a backup copy is made before beginning secondary editing. In addition to guarding against accidental data erasure, this backup allows survey personnel to review the changes made during secondary editing. It is recommend that the raw data be backed up to a secondary storage medium (e.g., a zip disk). If this not possible, make a backup copy of the raw data in a separate directory on the hard drive of the supervisor's machine.

Secondary editing

The goal of secondary editing is to investigate complex inconsistencies. The goal is not to systematically correct these inconsistencies. Inconsistencies should only by resolved if they concern essential variables (e.g., ages), if there is clear evidence of interviewer error or if the inconsistency is due to data entry error.

Secondary editing can involve several iterations because resolving one inconsistency often introduces others.

Final data backup

When secondary editing is complete, the final data are backed up. It is recommend that the final data be backed up to a secondary storage medium (e.g., a zip disk). If this is not possible, make a backup copy of the final data in a separate directory on the hard drive of the supervisor's machine.

The chart below illustrates the flow of secondary data processing.

```
        ┌─────────────────────┐
        │   Read Data into    │
        │       SPSS          │
        └─────────────────────┘
                   │
                   ▼
        ┌─────────────────────┐
        │   Recode Variables  │
        └─────────────────────┘
                   │
                   ▼
        ┌─────────────────────┐
        │    Add Sample       │
        │     Weights         │
        └─────────────────────┘
                   │
                   ▼
        ┌─────────────────────┐
        │     Generate        │
        │    Tabulations      │
        └─────────────────────┘
```
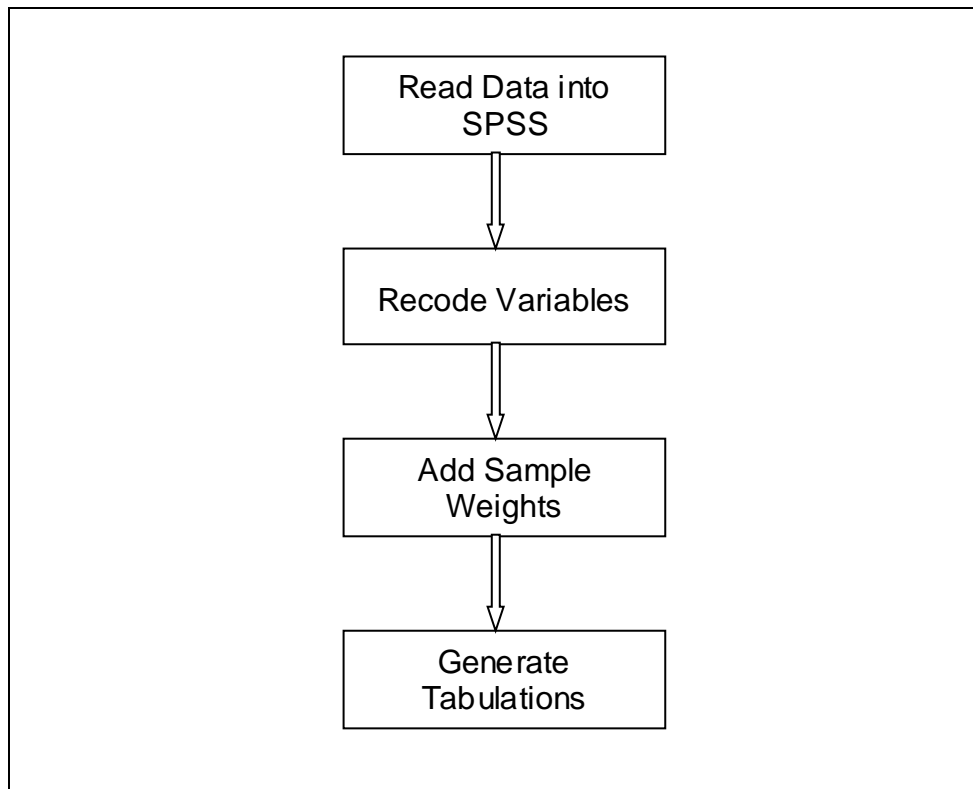
Concatenating data

The first step in secondary data processing is concatenating all of the cluster files into one national file.

Exporting data

The second step in secondary data processing is exporting the data from the data entry package into SPSS. This includes transferring any variable or value labels.

Recoding variables

Once the data have been exported to SPSS, the process of creating analysis files begins. The first step is to create new variables and to recode existing variables.

Add sample weights

Sample weights are calculated using an Excel spreadsheet. When all the required analysis variables have been created, the sample weights are added to the data files.

Generate tabulations

When analysis files have been created, SPSS programs generate the standard MICS tabulations.

Control Sheet

The control sheet is used to manage the flow of primary data processing. It is stored in an Excel file in the SUPER directory named CONTROL.XLS. The first column contains all of the

cluster numbers. Columns B, D, E, G, H, I and J contain space for writing the dates on which the primary data processing tasks are completed. Column C contains space for writing the number of the data entry operator who was assigned main data entry. Column F contains space for writing the number of the data entry operator who was assigned verification data entry.

Before the start of data entry, print a copy of CONTROL.XLS. This will serves as a paper backup of the control sheet. Each time a task is completed, update both the paper copy and the file. The second to last row of the control sheet contains the number of clusters for which each task has been completed. The last row of the control sheet contains the percent of clusters for which each task has been completed. Use these two rows to track the progress of the data processing system.

CONTROL.XLS needs to be modified for your survey. Modify the cluster numbers so that they match the cluster numbers for your survey, adding or deleting rows as necessary. Make sure that cells are correctly formatted (e.g., cells that are to contain dates are defined as date cells). Finally, check the formulas in the last two rows and update them if necessary.

Cluster summary sheet

A cluster summary sheet should be maintained for each cluster of questionnaires. It should be a large, heavy piece of paper with the cluster number written on it in large, bold numerals. The cluster summary sheet should be folded over the end of the package of questionnaires so that the cluster number is displayed clearly. The following fields should be printed on the top of the sheet.

Cluster Number   _____
Total Household Questionnaires   _____
Total Households Completed   _____
Total Women Questionnaires   _____
Total Children Questionnaires   _____

The questionnaire administrator should complete the fields when he or she receives the cluster.

Data entry directory structure

The diagram below shows the directory structure on a data entry machine.

```
MICS
      EPIINFO/IMPS/ISSA
            DATA
            ENTRY
            VERIFY
```

EpiInfo/Imps/Issa

This directory, which is named for the data entry package being used, contains the entry menu.

Data

The data directory contains the results of main data entry.

Entry

The entry directory contains the data entry programs.

Verify

The verify directory contains the results of verification data entry.

Supervisor directory structure

The diagram below shows the directory structure on the supervisor's machine.

```
MICS
        EPIINFO/IMPS/ISSA
                BACKUP
                EXPORT
                FINAL
                RAW
                SPSS
                SUPER
                WEIGHTS
```

EpiInfo/Imps/Issa

This directory, which is named for the data entry package being used, contains the supervisor menu.

Backup

The backup directory contains a backup of data files that have been structurally checked and verified but not edited.

Export

The export directory contains the programs used to export the data from the data entry package to SPSS.

Final

The final directory contains a backup of data files that have been structurally checked, verified and edited.

Raw

The raw directory contains the data files that have been transferred from the data entry machines.

Spss

The spss directory contains the programs that create analysis files and the tabulation programs.

Super

The super directory contains the programs that do structural checks, verification and secondary editing.

Weights

The weights directory contains the spreadsheet that calculates sample weights.

Data description

There is one file involved in creating a data description for ISSA. It contains the description of the structure of the data – levels or units of analysis (household or individual), records and variables. The records are defined as single (occurring one time) or multiple (occurring n times, i.e, the household listing). The filename has an extension of .DIC. The forms for data entry are also contained in this file.

There are three types of MICS questionnaires: household, women and children under five. The women's and children's questionnaires each correspond to a single unit of analysis: a woman and a child respectively. The household questionnaire contains two units of analysis: the household and the household members.

Each of the questionnaire types is stored in a separate file and each file uses a separate data description or dictionary. The data descriptions are named as follows.

Questionnaire	Data Description
Household	MICSHH
Women	MICSWM
Children	MICSCH

Modifying the data description

DIC files can be modified directly. With the dictionaries used for MICS, the data descriptions should be changed on the forms, adding or deleting variables as desired. It would be advisable to keep a backup of the original standard DIC files for reference.

Modules

The MICS questionnaires are made up of modules. Countries can add or remove modules, depending on their data needs. Countries are discouraged from significantly altering the contents of a module. The modules available for each questionnaire are listed below.

Questionnaire	Module	Code
Household	Household information	HI
	Household listing	HL
	Education	ED
	Child labor	CL
	Water and sanitation	WS
	Salt iodization	SI
Women	Women's Information panel	WI
	Child mortality	CM
	Tetanus toxoid	TT
	Maternal and newborn health	MN
	Contraceptive use	CU
	HIV/AIDS	HA
Children	Birth registration and early learning	BR
	Vitamin A	VA
	Breastfeeding	BF
	Care of illness	CI
	Malaria	ML
	Immunization	IM
	Anthropometry	AN

Modules in ISSA are located on separate records, since ISSA permits multiple record types. Each record is identified in ISSA by the two-character code listed above. Each record has one form or screen displaying all of the variables contained in that module.

Variable naming conventions

Variables are named for the module that they are in and the number of the question whose response they contain. For example, question 8 in the household listing is named HL8.

Some questions are split into two or more parts, with separate parts identified by a unique letter. Each part of the question is a separate variable. The names of these separate variables include the letters that distinguish the parts of the question. For example, question 3 of the women's information panel has two parts, 3A and 3B. Question 3B is named WI3B.

Some questions have two or more parts to the response categories. When these questions concern data, the letters D (for day), M (for month) and Y (for year) are appended to the base variable to create separate variables. In question 3A of the women's information panel, for example, both a month and a year are required. These two variables are named WI3AM and WI3AY.

There are two questions in the questionnaires in which the first part of the response is the form of the response and the second part is the response: question 6 in the maternal and newborn health module of the women's questionnaire and question 3 of the anthropometry module of the children's questionnaire. In both cases, the variable giving the form of the response has the letter A appended to its name while the variable containing the response has no letter appended (e.g., MN6A and MN6, AN2A and AN2).

There are a number of questions in the questionnaire that allow for multiple responses. For these questions, each response has its own variable. The variables' names are made up of the module code, the question number and the letters A through K, which correspond to the response codes 1 to 11. An exception is the response "don't know" which uses the letter Z in all cases. For example, question 4 of the malaria module in the children's questionnaire has variables named ML4A, ML4B, ML4C, ML4D, ML4Z.

Identification variables

Every questionnaire must have a series of variables that uniquely identifies it. The variables that identify a questionnaire are known as the identification variables. The table below lists the questionnaire types and their identification variables.

Questionnaire type	Cluster number	Household number	Line number
Household	HI1	HI2	
Women	WICLNO	WIHHNO	WILNNO
Children	CHCLNO	CHHHNO	CHLNNO

MICSHH.DIC

The household DIC is unremarkable except for one added variables: HIMEM. This variable does not exist on the MICS standard questionnaires. It has been added to increase control over data entry. HIMEM is used in the data entry program to ensure that the correct number of household members is entered in the household listing module.

It should be noted that a household questionnaire containing only the information panel and the household listing should have a third record type added that includes a variable INDIV (values Y or N) to help to control the entry of women and/or children records. They are being stored not as a lower level in ISSA but in separate external files. During data modification,

this single variable will provide a field to easily advance to (F6) for modifying any women or children data.

MICSWM.DIC

The women's DIC file contains only the variables that exist in the questionnaire. Multiple response variables exist in the MICSWM.DIC. They are MN2A through MN2Z, MN3A through MN3Z, and CU4A through CU4N. See **Multiple response questions** below.

MICSCH.DIC

The children's DIC file, like the woman's DIC file, contains only variables that exist in the questionnaire. There are multiple response variables in the MICSCH.DIC. They are CI10A through CI10K, CI11A through CI11J, ML4A through ML4Z, and ML7A through ML7Z. See **Multiple response questions** below.

In the anthropometry module, the first 5 variables will be entered by the data entry operator. Variables HAP through FLAG will have values assigned by the data entry program functions described below.

Data entry programs

ISSA's data entry programs are known as applications. Their filenames have an APP extension. They are used to check ranges with more than 3 sets of values, check basic consistency, control the data entry path (the path that a data entry operator follows through the questionnaire) and link the external files to the household file.

Coding conventions

The standard programs use standard coding for certain responses. The response "don't know" is always coded as a nine with leading nines. Questions that are not applicable to a respondent are always coded as an eight with leading nines. Questions with a missing response (i.e., the interviewer did not record a response to an applicable question) are always coded as a seven with leading nines. Responses that are inconsistent with other data in the questionnaire are always coded as a six with leading nines. The table below summarizes the standard coding conventions.

Response	Variable length			
	One character	Two characters	Three characters	Four characters
Don't know	9	99	999	9999
Not applicable	8	98	998	9998
Missing	7	97	997	9997
Inconsistent	6	96	996	9996

All the response categories for a variable should contain the same number of digits. If a question requires two digits, codes with values less than 10 must have leading zeros. For example, if a question uses codes 1 through 12, the codes on the questionnaire must be 01-12 and not 1-12.

Because the codes 7, 8 and 9 are reserved for special use, any question that requires more than 6 response categories must have 2-digit responses categories with leading zeros (e.g., 01, 02, 03, 04, 05, 06, 07, 97 and 99).

Ranges

Most of the variables in the MICS questionnaires have defined ranges. Only values within the specified ranges can be entered. This reduces the probability of data entry errors. The only variables without defined ranges are variables for which it is impossible to create restrictive ranges. The limits of household number, for example, vary so much from cluster to cluster that any defined ranges would catch only a small number of errors.

ISSA allows an infinite number of ranges. Ranges that can be defined within 3 sets of values are checked by the dictionary range definition. Variables with more than 3 sets of range values must be checked in the application, usually with a BOX-ENDBOX , ERRMSG combination of logic.

Skips

The MICS questionnaires make abundant use of skips. Skips are instructions on the questionnaire that tell the interviewer to skip all the questions between the current question and a question later on in the questionnaire. Skips on a questionnaire must be matched by skips in the corresponding data entry program. Skips in a data entry program help define the data entry path.

ISSA strictly enforces the data entry path as long as the command NOINPUT is not used. Only in that case can a data entry operator enter an erroneous value. If NOINPUT is used in

programming, it is necessary to use a POSTPROC error check to make certain the data entry operator has not changed a value defined by the program by "backing into" the variable.

Error messages

Error messages are displayed when the ERRMSG function is called. The error messages are numbered and a master list of the error message number exists. If you add error messages, make sure that you do not reuse a number.

If you add error messages, it is important to add them to the editing document used by the data entry supervisor and secondary editors so that they will know what step is to be taken if the error message is produced.

Many of the error messages are followed by a REENTER command that returns to the field that is being entered. This forces the entry operator to address the error before advancing. Because the data entry operator will at times be required to enter corrections, careful supervision is necessary.

Multiple response questions

There are a number of questions in the questionnaire that allow for multiple responses. For these questions, each response has its own variable. Each variable has three possible responses – 0 for no, the responses code on the questionnaire and 7 for missing data.

For example, the variable MN2 records whom a woman saw for antenatal care before her last birth. The response code 2 is associated with the response nurse/midwife. The variable for this response is named MN2B. The allowed entries are 0, 2 or 7.

If a missing value is entered for any of the responses, a missing value must be entered for all of the responses. The number of missing values must therefore equal zero or the number of responses.

This requirement is checked using the working variable MISS. MISS is initialized to zero and then incremented by one if a missing value is entered for a response. When the last response has been entered, MISS must equal either zero or the number of responses. If it does not, an error message is displayed and the entry operator is unable to advance until the inconsistency has been resolved.

MICSDE.APP

There is only one data entry program in ISSA. The input dictionary is the MICSHH dictionary. At the conclusion of the entry of the household, the household listing is checked for eligible women and/or children. If eligible women are found, the external dictionary MICSWM is opened and data for women are entered. After all women have been entered, the external dictionary MICSCH is opened and data for children are entered.

The logic for opening the external dictionaries is found in the POSTPROC for the section MODSI. If this module is not used in a country, it is imperative that the logic found in the MODSI POSTPROC be placed in the last single section of the MICSHH dictionary as the POSTPROC. As noted above, if no single section follows the household listing, a single section with one variable in it should be added to the dictionary for ease in updating the women and children files.

Throughout this section we will refer to blocks of code related to a specific level, section or variable as procedures. Thus, the code related to the variable HI1 is the cluster number procedure. Only procedures that contain unusual or complex code will be discussed.

LEVELID

The only LEVELID declaration found in the data entry program is for level 1, the household. This LEVELID must include the variables required to uniquely identify a household.

The identification logic for the other units of analysis (women and children) is found in the POSTPROC for MODSI. The woman's identification variables are copied from the household identification variables with her household listing line number (HL1) completing the ID. The same variables are used for the children's file.

User defined functions

The first 7 user-defined functions (ZSCOEF, DABS, ZSPCT, ZSEVAL, ZSCR, ZSANTH, and AGEMTH) are used in the child's questionnaire to calculate the anthropometry scores that are found at the end the anthropometry module. The AGEMTH function is called to calculate the child's age in months. The ZSANTH function is then called. This function calls ZSEVAL, SZCR and ZSPCT. The function ZSEVAL calls ZSCOEF, and ZSPCT calls DABS.

The code in the AGEMTH function calculates the age of the child in months. Because anthropometry is highly sensitive to age, the age of the child must be based on the child's age in days. The code first calculates the number of days that elapsed between the beginning of the year and a child's birth. It then calculates the number of days that elapsed between the beginning of the year and the date of interview. Finally the number of days in the years between the year of birth and the year of interview is added to the number of days since the beginning of the year until the date of interview. The difference between these two numbers of days is the child's age in days. This is then converted into the child's age in months by dividing by 30.4375 (the average number of days in a month over four years). Because of the need for accuracy, the child' age in months is calculated to two decimal places.

The VALID function is used to check for special values that represent not applicable, missing, don't know or inconsistent values. The NATOZERO function is used to change a not applicable value to 0, allowing it to be added to another variable. See one use of it in the MODCM variable CM9. The function VCHECK is used to check the values entered into the vaccination dates in the MODIM module.

Date related ISSA functions

ISSA uses century month codes (CMC) for all dates in the data entry program. A CMC is the number of months since the turn of the 20th century. For example, the CMC for January 1900 is 1; the CMC for March 2000 is 1203. CMCs are calculated as follows: subtract 1900 from the year of the event; multiply that number of years by 12, the number of months in a year. Add the month value to that product.

The four CMC functions used in the data entry program are used to set the CMC for a particular date and if an age is related to that date, adjust the CMC using the age. The following table is a brief explanation of the functions found in the MICSDE application.

Lcmc = setlb (month, year, mimimum);	The month, year and minimum CMC are passed into the function. If the year is not valid, the minimum is returned. If the month is not valid, the CMC for January of the year is returned. If both year and month are valid, the CMC is calculated and returned.
Ucmc = setub (month, year, maximum);	The month, year and maximum CMC are passed into the function. If the year is not valid, the maximum is returned. If the month is not valid, the CMC for December of the year is returned. If both year and month are valid, the CMC is calculated and returned.

T = adjjlba (lcmc, ucmc, di, di, age);
If t < 0 then
   E = errmsg (2171);
Else
   Lcmc = t;
Endif;
	The lowest possible CMC for a date, the highest possible CMC for a date, the CMC for the date of interview, and the age are passed into the function. The age is used to attempt to raise the lcmc to the minimum allowed for that age, given the date of interview, while keeping it below or equal to the ucmc. If the age would cause the lcmc to be greater than the ucmc (age is too young for the date of birth), the function returns −1, signaling an error. If the age would cause the lcmc to be less than the original lcmc (age is too old), the function returns the original lcmc. This error for age is caught by the use of the function adjuba. If the age would cause the lcmc to be greater than it is and less than or equal to the ucmc, the function returns that value and the lcmc should be raised to that value.

T = adjuba (lcmc, ucmc, di, di, age);
If t < 0 then
   E = errmsg (2171);
Else
   Ucmc = t;
Endif;
	The lowest possible CMC for a date, the highest possible CMC for a date, the CMC for the date of interview, and the age are passed into the function. The age is used to attempt to lower the ucmc to the maximum allowed for that age, given the date of interview, while keeping it above or equal to the lcmc. If the age would cause the ucmc to be less than the lcmc (age is too old for the date of birth), the function returns −1, signaling an error. If the age would cause the ucmc to be greater than the original ucmc (age is too young), the function returns the original ucmc. This error is caught by the use of the function adjlba. If the age would cause the ucmc to be less than it is and greater than or equal to the lcmc, the function returns that value and the ucmc should be lowered to that value.

PREPROC procedure

Unless specified as a PREPROC, all procedures in ISSA are assumed to be POSTPROCs. The PREPROC at LEVEL 0 is used to set constants that are used throughout the data entry program.

HI1 procedure

When the data entry program is launched, a variable is set for the cluster number found in the data file name. This variable is then used to define the legal value for the HI1 field. This ensures that the data file name and the cluster identification in the data file are consistent.

MODSI procedure

The logic found here allows the entry and modification of women and child questionnaires. The operator will be prompted for each eligible woman and then child questionnaire. If a questionnaire is not present because it was not completed, the operator has the option of skipping the questionnaire. For both the women and child data, the logic is the same:

1. Clear the memory that stores the individual's data (x=clear(MICSaa);)
2. Check to see if the individual questionnaire exists
   (I = loadcase (MICSaa,HI1,HI2,HL1(ln);)
3. If the case exists (I = 1) then the options presented to the operator are to modify or skip the questionnaire.
4. If the case does not exist (I = 0) then the options presented to the operator are to enter or skip the questionnaire.
5. Finally, if the individual is being entered for the first time or the operator responds "Yes" to the prompt of "Update questionnaire", the data are written to the external file.

Structure checking

Structure checking is an essential part of the MICS data processing approach. It is essential that the data be structurally sound. The data entry program enforces some structural coherence, but not all. There are several reason for this, but the most important is that it would constrain data entry too much. Therefore, a structural checking program must do so.

Using a cluster control sheet with each cluster of questionnaires will help to resolve problems found in the structure checking program. This cluster control sheet should be a piece of heavy paper that folds over the end of the package of questionnaires to display the cluster number in bold numbers. On the top of the sheet, the following fields for questionnaire totals should be printed:

Cluster Number   _____

      Total Household Questionnaires   _____

      Total Households Completed   _____

      Total Women Questionnaires   _____

      Total Children Questionnaires   _____

___

101

When the cluster is checked by the Questionnaire Administrator, these totals should be written on the sheet.

Assumptions

Before examining the structure checking program, consider the lists below that lay out some of the structural assumptions that are being checked. The lists are not exhaustive, but they do illustrate the general principles on which structural checks are based.

Household

- The number of households should equal the sum of the number of incomplete and complete households.
- The number of records in the household data file should equal the number of households implied by the household listing data file.

Household listing

- The number of household members listed on the household cover sheet should equal the number of household member records in the household listing data file.

- The number of household members who are women age 15 to 49 should match the number of eligible women on the household cover sheet.
- The number of household members who are children under the age of 5 should match the number of eligible children on the household cover sheet.

Women

- The number of women's records should match the number of interviewed women on the household cover sheet.
- There should be a one-to-one correspondence between records in the women's data files and records in the household listing data file for women age 15 to 49.

Children

- The number of children's records should match the number of completed children interviews on the household cover sheet.
- There should be a one-to-one correspondence between records in the children's data files and records in the household listing data file for children under the age of 5.

Notice that many of the assumptions above are true for both the household and the cluster. For example, the number of eligible women on the household cover sheet should equal the number of female household members age 15 to 49. At the same time, the sum of eligible women for all households in the cluster should equal the number of female household members age 15 to 49 for the entire cluster.

PROGRAMS

Structure checking is done using CHECK.APP in the SUPER directory. This program checks the structural coherence of the household, women and children data files.

The focus here is on what the programs are doing instead of how they do it. The programs are complex and do not lend themselves to facile explanations. The best way to understand them is to study them carefully to understand what they are doing.

The output from the CHECK.APP is in 2 formats. The first type of output is a tabular summary of what was found in the 3 types of questionnaires, whether inconsistent or not. Information from the Household Information Module is presented along with the count of eligible women and children expected and found in the household. A member by member list is presented showing their eligibility for individual questionnaires. Finally, a summary of the entire cluster is presented.

The output below shows an example based on one household.

```
                        MICS Data Structure Check
        Cluster:    3

            Households       |Household Members|     Women         |     Children
        Total  Comp  Incomp|  Cover      Found | Elig. Int. Found | Elig. Int. Found
          1     1     0  |    6          5  |   2    2    2  |   0    0    1

        ** Total household members incorrect
        ** Total children interviewed incorrect
```

There are two structural problems in this example. First, the household cover sheet records 6 household members but only 5 household listing records exist in the data file. Second, there is a case in the child's data file, but the household cover sheet records no children under the age of 5 and no children's interviews.

```
Household       HI10 | HIMEM          |HI11    HI12     |HI13    HI14
Household member     |          HL1   | HL5     WILNNO  | HL7      CHLNNO

     3001   res=1 |    6           | 2      2        | 0        0
     300101       |          1     | 0          .    | 0          .
     300102       |          2     | 2          2    | 0          .
     300103       |          3     | 3          .    | 0          .
     300104       |          4     | 0          .    | 0          .
     300105       |          5     | 0          .    | 0          5
** 0210 Interviewed child not eligible
** Number of household members incorrect
** 0209 Number of interviewed women incorrect
** 0208 Number of interviewed children incorrect
```

HI10 is the result of the household interview. In the example, the household interview was completed. HIMEM is the number of household members. Notice that it is followed by only 5 household members; the household member will line number 6 is missing.

HI11 is the number of eligible women. HL5 is a household variable that reflects eligibility for the women's interview. Two household members are recorded as eligible, a total that matches HL5. HI12 is the number of interviewed women. WILNNO is the line number for a record in the women's data file. Only one household member has a value meaning that one women's questionnaire is missing for this household, one women's questionnaire for this cluster has an incorrect line number or the value of HI12 is incorrect.

HI13 is the number of eligible children; HL7 is the line number of an eligible child's caretaker. HI14 is the number of complete children's interviews. CHLNNO is the line number for a record in the children's data file. One such record exists despite the value of HI14 (0) meaning that the eligibility code of one of the household members is incorrect and the values HI13 and HI14 are incorrect or the household identification for the child's questionnaire is incorrect.

```
      Households      |Household Members|     Women          |    Children
Total  Comp  Incomp|  Cover    Listed | Elig.  Int.  Found | Elig.  Int.  Found
  1     1      0   |    6        5    |   2     2     1    |   0     0     1

** Total household members incorrect
** Total women interviewed incorrect
** Total children interviewed incorrect

** Total women's questionnaires inconsistent
```

The missing women's questionnaire had an incorrect line number: 9 instead of 3. Because no household listing record has a line number of 9, this women's record could not be matched to a household listing record and was not counted. The error messages that follow the table summarize the errors that can be observed in the table.

In addition to the tabular summaries produced by CHECK.APP, messages are produced that identify inconsistencies within questionnaires, such as missing or inappropriate records for a case. These messages also will flag households for which the presence or absence of eligible women or children questionnaires appears inconsistent.

All of these errors should be carefully checked against the bundle of questionnaires for the cluster. It is necessary to correct all errors found by this application.

Verification

Verification is done using an ISSA utility EXCOMPAR.EXE. It is possible to exclude certain variables from the comparison, by creating an application listing the variables to be excluded. Generally, only the date entry operator variable should be excluded. Other possibile exclusions would be alphabetic variables that contain the names of persons in the household.

The three applications used for verification are COMPHH.APP, COMPWM.APP and COMPCH.APP. Each application contains a list of variables that will not be compared during verification. It is recommended that NO additional variables be added to these lists since differences in other variables can affect the quality of the data.

The main data entry file (which has been copied onto the supervisor's machine) is compared to verification data entry file (which has been copied onto a diskette). The verification utility will produce a list of differences if any exist. If there are differences, this list should be printed and given to the two data entry operators.

The data entry operators then use the list of differences and the cluster's questionnaires to check each difference and record on the list which file needs to be corrected. When all of the differences have been investigated, the data entry operators correct any errors in their files. They then recopy the data files to their floppies and the files are compared again on the supervisor's machine. This process continues until the files are identical.

Secondary editing

The logic found in the secondary editing is relatively straight forward.   The functions are the same as those found in the data entry application. Again, the editing manual must be the continual reference for making corrections to data that appear inconsistent.

In the vaccination section, there is logic to check the dates of vaccinations with the date of birth of the child and the date of interview. In addition, there are consistency checks done among vaccination dates for vaccinations that are given in series.

It should be remembered that any changes to the standard MICS questionnaires should be reviewed for additional editing checks that must be incorporated into the entry and/or editing programs.

Menu systems

Two menus manage the MICS data processing system. Data entry operators use the data entry menu to enter, modify and transfer data. The data processing supervisor uses the supervisor menu to execute the structure checking program, transfer and verify data, backup the raw data, execute the editing program, modify data during secondary editing and backup data files.

Both menus are implemented using DOS batch files.

Data entry menu

The diagram below shows the contents and structure of the data entry menu. Typing ENTRY in the C:\MICS\ISSA directory launches the menu system.

1. **Select a New Cluster Number**

2. **Enter Main Data**
3. **Modify Main Data**

4. **Enter Verification Data**
5. **Modify Verification Data**

6. **Index Main Household Data**
7. **Index Main Woman's Data**
8. **Index Main Child's Data**

9. **Index Verification Household Data**
10. **Index Verification Woman's Data**
11. **Index Verification Child's Data**

12. **Upload Main Data Files**
13. **Upload Verification Data Files**
14. **Quit**

1. Select a New Cluster Number

This choice permits the data entry operator to select a new cluster number for main data entry or verification.

The cluster number is checked as data entry begins to make certain that it is an acceptable value for a cluster number.

2. Enter Main Data

This option permits the data entry operator to add new households (and women/children) to a Main Data File. This option is chosen when the data entry operator begins a new cluster or when he/she continues data entry of a Main Data File after a break. The data are stored in the directory MICS\ISSA\DATA.

3. Modify Main Data

This option permits the data entry operator to modify data already entered into a Main Data File. If the file has not yet been created, the operator will be returned to the menu. The data are stored in the directory MICS\ISSA\DATA.

This same option is available to the data entry supervisor on his/her supervisor menu. However, the data on the supervisor's machine are stored in MICS\ISSA\RAW.

An important point to stress in training the data entry operators is the method of accessing the women and children individual data. This is done by selecting the household and while editing the household ADVANCE (F6) to the last field in the household. In the standard MICS application, that field is SI1. From there the operator will be prompted to modify or skip each individual who may or may not have an individual questionnaire.

4. Enter Verification Data

This option permits the data entry operator to add new households (and women/children) to a Verification Data File. This option is chosen when the data entry operator begins a new cluster or when he/she continues data entry of a Verification Data File after a break. The data are stored in the directory MICS\ISSA\VERIFY.

5. Modify Verification Data

This option permits the data entry operator to modify data already entered into a Verification Data File. If the file has not yet been created, the operator will be returned to the menu. The data are stored in the directory MICS\ISSA\VERIFY.

6-8. Index Main Household (Woman's, Child's) Data File

These three options allow the data entry operator to reindex the Main Data Files if an index is destroyed. These options are seldom necessary.

9-11.Index Verification Household (Woman's, Child's) Data File

These three options allow the data entry operator to reindex the Verification Data Files if an index is destroyed. These options are seldom necessary.

12. Upload Main Data Files

This option allows the data entry operator to copy the Main Data Files from his/her data entry machine to a floppy for supervisory activities on the supervisor's machine. After main data entry, the floppy is read by the supervisor's machine to check the structure. For verification, backup and secondary editing activities, the Main Data Files are copied to the supervisor's machine by the supervisor's menu.

13. Upload Verification Data Files

This option allows the data entry operator to copy the Verification Data Files from his/her data entry machine to a floppy for supervisory activities on the supervisor's machine. After verification data entry, the floppy is read by the supervisor's machine to check for differences between the Verification Data File and the Main Data File.

Supervisor menu

The diagram below shows the contents and structure of the supervisor menu. Typing SUPER in the C:\MICS\ISSA directory launches the menu system.

1. Check Data Structure
2. Download Complete Data File and Verify
3. Save Raw Data Backup

4. Run Secondary Edits
5. Modify Data

6. Save Final Data Backup

7. Select New Cluster

8. Quit

1. Check Data Structure

This option reads the Main Data File from the floppy drive and checks the structure of the data files for households, women and children. After execution of the CHECK.APP, the supervisor will be presented the output on the screen and can print it IF errors need to be corrected. Any errors in this output should be corrected by carefully checking the cluster of questionnaires.

2. Download Complete Data File and Verify

This option should be chosen only after no errors exist in the data from the data structure check program. This option will copy the Main Data Files from the floppy to the C:\MICS\ISSA\RAW directory and will read the Verification Data Files from the floppy diskette inserted AFTER the Main Data Files have been copied.

The output from this option will list any differences between the data files produced by Main Data Entry and Verification Entry operators. The output will be presented to the supervisor and he/she will be able to print the listing IF differences exist. All differences must be resolved by the data entry operators and corrections made to their data files. After corrections, they should again chose to upload the data files to floppy and the supervisor should again use the Verify Option until no differences exist between the main and verification files.

When differences exist, the Main Data File is erased from the MICS\ISSA\RAW directory. Only if no differences exist can the processing continue to the next step.

3. Save Raw Data Backup

This option permits the supervisor to backup the raw data file after structure checks and verification checks have been passed. This should be done before any secondary editing has been done to the data. The data will be stored in the C:\MICS\ISSA\BACKUP directory.

4. Run Secondary Edits

This option allows the supervisor to execute the EDITING.APP application on the verified raw data. The output from this program will be presented to the supervisor and should be printed for review by the secondary editors. The Editing Manual should be used during all editing.

This option should be selected again AFTER the supervisor has made any corrections to the data file based on corrections listed by secondary editors.

5. Modify Data

This option permits the supervisor to use the data entry application to correct the data stored in the C:\MICS\ISSA\RAW directory. The supervisor should use this option rather than the

ENTRY menu since the ENTRY menu is designed for the directory structures on the data entry machines.

After using this option, the supervisor should return to the 4th option (Run Secondary Edits) to make certain that changes to the data have corrected the errors as desired and that no new errors have been created.

6. Save Final Data Backup

This option allows the supervisor to copy the final data files to the C:\MICS\ISSA\FINAL data directory. Data stored in this directory will later be concatenated, sorted and exported into SPSS data files.

7. Select New Cluster

This option allows the supervisor to begin processing a new cluster number.

Transferring data

Generating the MICS standard tabulations involves three distinct steps: transferring the data from ISSA to SPSS, creating analysis files and running tabulations. Standard programs exist for each step, but these programs must be carefully checked and adapted if necessary.

Concatenation

When data entry is complete, three final (i.e., verified and edited) data files will exist for each cluster: one for households, one for women and one for children. These cluster data files must be concatenate into three national files. The files can be concatenated using the CONCAT.BAT batch file. To use CONCAT.BAT

1. Open an MS-DOS window and navigate to the EXPORT sub-directory (e.g., \MICS\ISSA\EXPORT)
2. Type CONCAT and press the enter key

The CONCAT.BAT file will create three new files CH.DAT, HH.DAT and WM.DAT. CONCAT.BAT will also sort these new data files using three ISSA sorting applications, SORTHH.APP, SORTWM.APP and SORTCH.APP. Consider the syntax listed below which executes SORTHH.ATT.

exsort sorthh micshh=hha.dat $copy=hh.dat

Exsort is an ISSA program that sorts data files. The first parameter is the name of the sort application which defines the variables on which the data are to be sorted (HI1 and HI2). The second parameter specifies the dictionary (MICSHH.DIC) and the data file (HHA.DAT) that it describes. The third parameter specifies the name of the sorted output file.

Transferring the data

CONCAT.BAT will also transfer the data files from ISSA to SPSS. The data are transferred by four ISSA export applications: READHH.APP, READHL.APP, READWM.APP and READCH.APP. After the data have been transferred, the ISSA export description creates an SPSS syntax file that describes them.

The statement below exports the household data file to an ASCII data format that can be read into SPSS.

```
exexport readhh micshh=hh.dat $copy=readhh.dat
```

The first parameter following the exexport program call is the name of the export application, READHH.APP. The second parameter specifies the dictionary (MICSHH.DIC) and the data file (HH.DAT) that it describes. The third parameter specifies the name of the output data file, READHH.DAT.

The statement below creates the SPSS syntax file that describes READHH.DAT.

```
exexpdes readhh spss
```

The first parameter following the exexpdes program call is the name of the data file. The second parameter is the package for which a data description file must be created.

Output

When the CONCAT.BAT is executed, it will produce four ASCII data files (READHH.DAT, READHL.DAT, READWM.DAT and READCH.DAT) and four SPSS data description files (READHH.SPS, READHL.SPS, READWM.SPS and READCH.SPS – collectively known as the READ files).

The SPSS command

```
save outfile = 'filename.sav'.
```

must be modified and added to the end of each data description file. The word FILENAME must be replace with HH, HL, WM or CH, depending on the type of data file. This command will save the data file that is created by each syntax file.

Once this command has been added to each data description file, executing the READ files will create the SPSS data files HH.SAV, HL.SAV, WM.SAV and CH.SAV.

Creating Analysis Files

Adding labels

Before using a data file, it is helpful to add variable and value labels. A variable label describes the subject matter of a variable. Value labels describe what each value of a variable represents.

Consider the household listing variable HL3 that stores the sex of household members. It takes a value of 1 for males, 2 for females, and 7 for missing data. The variable should be labeled "Sex". The value 1 should be labeled "Male". The value 2 should be labeled "Female". The value 7 should be labeled "Missing".

The SPSS programs LABELHH.SPS, LABELHL.SPS, LABELWM.SPS and LABELCH.SPS (collectively known as the LABEL programs) assign variable and value labels to all the variables in the standard questionnaire. If you have added variables to the questionnaire, go through these programs and add variable and value labels where appropriate.

The LABEL programs are unnecessary if you are using ISSA; ISSA transfers variable and value labels when it exports data.

Adding variables from other files

The SPSS program MAKEHL.SPS, MAKEWM.SPS and MAKECH.SPS (collectively known as the MAKE programs) perform two functions: they add variables from other files and they recode variables to create new variables. Each of these functions will be dealt with in separate sections. This section will concern itself with adding variables from other files.

When analyzing household members, it is useful to have access to variables on the household questionnaire (e.g., HI6, HI7). These variables can be added to the household listing data file by merging it the household data file. More broadly, useful variables can be added to the household listing, women's and children's data files from other MICS data files. The table below shows these three data files and the files with which they are merged.

Base file	External file
HL.SAV	HH.SAV
WM.SAV	HL.SAV
CH.SAV	WM.SAV

These merges are executed in the order implied by tables. As a result, when the household listing file is added to the women's file, it adds both household and household listing variables. Similarly, when the women's file is added to the children's file, it adds household, household listing and women's variables.

SPSS Macros

Macros in SPSS are used to automate a series of statements that are used repeatedly. The MAKE programs use macros to simplify the process of merging files. MAKEWM.SPS and MAKECH use the same macro.

The code below is the macro used in MAKEWM.SPS and MAKECH.SPS.

```
*define a MACRO that merges 2 files with 3 ID variables.
define !merge (file1 = !tokens(1)
               /mvar1 = !tokens(1)
               /mvar2 = !tokens(1)
               /mvar3 = !tokens(1)
               /file2 = !tokens(1)
               /name1 = !tokens(1)
               /name2 = !tokens(1)
               /name3 = !tokens(1)).
get file = !file1.
sort cases by !mvar1 !mvar2 !mvar3.
save outfile = 'tmp.sav'
  /rename = (!mvar1 = !name1) (!mvar2 = !name2) (!mvar3 =
!name3).
get file = !file2.
sort cases by !name1 !name2 !name3.
match files
  /file = *
  /table = 'tmp.sav'
  /by !name1 !name2 !name3.
save outfile = !file2.
get file = !file2.
erase file = 'tmp.sav'.
!enddefine.
```

The define keyword begins the macro and the enddefine keyword ends it. The macro's name, merge, immediately follows the define keyword and is preceded by the exclamation point character.

The text enclosed in parenthesis that follows the macro's name defines accepted parameters. For example, file1 and file2 are parameters that represent file names. When a parameter is referred to in the main body of the macro, it is preceded by the exclamation point. The remaining statements in the macro are standard SPSS commands that will not be discussed here.

The statements above define the macro but do not execute any of the commands within. The commands within are only executed when the macro is called.

A macro is called by typing its name preceded by the exclamation point. When it is called, any parameters used within the macro must be specified. The following lines show the calling of the MERGE macro in the MAKEWM.SPS program.

```
*add variables from the household listing file.
!merge file1=hl.sav mvar1=HI1 mvar2=HI2 mvar3=HL1 file2=wm.sav
name1=WICLNO name2=WIHHNO name3=WILNNO.
```

This statement will merge the files HL.SAV and WM.SAV. The files will be merged using the cluster number, the household number and the woman's line number.

Warning!

The merge above is of a keyed table (e.g., HL.SAV) onto a file (WM.SAV). With this type of merge, extra cases in the keyed table (e.g., household members who are not women aged 15-49) are not added to the file. In other types of merges, these extra cases would be appended to the end of the file. If you modify the match files statement, do not replace the /table specification with a /file specification. This will result in ineligible household members being added to the women's data file. This warning applies to all of the MAKE programs, not just the MAKEWM.SPS program.

Recoding variables

Variables used in the standard tabulation plan are recoded in two places: the MAKE programs and the tabulation programs. Variables that are used in several tabulations are recoded in the MAKE programs. All other variables are created in the tabulation programs.

The recoding of most variables uses standard SPSS commands and will not be discussed here. There are however a few unusual approaches that will be explained.

One of these is the recoding of variables into 0 or 100. This unusual recoding is down for presentation purposes only. When SPSS displays percentages in a table, it displays all categories. For many tables in the tabulation plan, we are only interested in one category. If we assign a value of 100 to that category and a value of 0 to all other categories, the mean of the variable is the percentage of respondents in that category.

For example, the variable HA1 records whether a woman has heard of AIDS. It takes a value of 1 if a woman has heard of AIDS and a value of 2 if she has not. In the first column of table 30, we want to display only the percentage of women who have heard of AIDS.

In the program MAKEWM.SPS, the variable HA1 is recoded into the variable AIDS. AIDS takes a value of 100 if the woman has heard of AIDS and a value of 0 otherwise. The mean of AIDS is the percent of women who have heard of AIDS. To understand why this is so, consider the example below.

Women who have heard of AIDS	10
Total number of women	20
Percentage of women who have heard of AIDS	**10 / 20 * 100 or 50%**
Mean of the variable AIDS	(10*100 + 10*0) / 20
	10 * 100 / 20
	10 / 20 * 100 or 50%

Adding Weights

Sample weights must be used when tabulating MICS data unless your country used a self-weighting sample. There are three kinds of weights: household, women's and children's. Household weights should be used with the household and household listing data files. Women's weights should be used with the women's data file. Children's weights should be used with the children's data file.

Weights must be used if the probability of selecting a household for interview was not equal across all of your country. For example, consider the imaginary country Popsylvania. In the North region, 5 households per 10,000 were selected for interview. In the capital, Vladisville, 28 households per 10,000 were selected for interview. In the country as a whole, 10 households per 10,000 were selected for interview.

Sample weights must be used to adjust the sample to produce accurate estimates for the whole country. The sample weights used are the inverse of the relative probabilities of selection. For example, the sample weights for the North region and Vladisville are

North Region	10/5	2.00
Vladisville	10/28	0.35

The weights for your country are calculated in a spreadsheet named WEIGHTS.XLS. The first step is to run the SPSS program WGTTAB.SPS. This program will produce the information needed to complete the WEIGHTS.XLS spreadsheet. Add this information to WEIGHTS.XLS and it will automatically calculate sample weights in columns Q, W and AC.

Save WEIGHTS.XLS when you have made these changes. Then close Excel and open SPSS. From the file menu, select the open option. Select WEIGHTS.XLS and click the open button. SPSS will pop-up the opening files options dialog box. Enter in this box the range of cells in the spreadsheet that you want to read. In the standard version, the range of cells is A5:AD16.

When you have entered the range of cells, click ok. SPSS will read in the specified range of cells and assign variable names to each column based on the values in row 5. In addition to the rows of the spreadsheet that you need, SPSS will also read in three rows that you do not need. Delete rows 1, 6 and 7 from your SPSS data set. Save the file in the SPSS sub-directory under the name WEIGHTS.SAV when these modifications have been made.

WEIGHTS.SPS merges the appropriate weights in WEIGHTS.SAV onto HH.SAV, HL.SAV, WM.SAV and CH.SAV using a macro. WEIGHTS.SPS also deletes unnecessary variables in WEIGHTS.SAV before it merges it with the data files. WEIGHTS.SPS does not need to be modified if your sample is stratified based on area and region.

Output

When labels, variables and weights have been added to these SPSS files and variables have been recoded, you will have four analysis files. The tabulation programs will use these four data files to create the tables in the tabulation plan.

Tabulation

There is one tabulation program for each table. Each program's name is the letter T followed by the number of the table in the tabulation plan. For example, the program T1.SPS creates table 1 in the tabulation plan.

Each tabulation program must be carefully reviewed. One important thing to check is whether the variables used in the tabulation program exist in your data file. If they do not, check whether the variable is of primary or secondary importance. If a variable of primary importance does not exist in your data file, you must either remove the table entirely or ask an analyst to redesign the table. If a variable of secondary importance is missing, remove all reference to the variable in the tabulation program and make any other adjustments necessitated by its absence.

All recoding activity must also be carefully checked. If there are variables on your questionnaire that have non-standard categories, any recoding activity involving those variables must be examined. If your questionnaire contains non-standard variables, they must be recoded if they are to be included in tabulations.

Weighting in the tabulation programs is straightforward except where the SPSS aggregate command is involved. If the goal of the aggregate command is to cumulate across cases to calculate a numerator and a denominator, weights must be applied before the aggregate command. They should not be used when working with the resulting file; it has already been weighted.

For example, table 1 contains the household responses rate. The household response rate is difficult to calculate because it requires dividing one variable by another within the table. One solution to this problem is to create an aggregate file that contains counts of sampled households, occupied households and interviewed households. The aggregate file will contain one case for each category of the specified break variable.

The weights must be applied when the aggregate file is created to generate the weighted numerator (the count of interviewed households) and the weighted denominator (the count of occupied households). Once the aggregate file has been created, the household response rate for each category of the break variable is the numerator divided by the denominator.

If the goal of the aggregate command is to create a summary statistic for individual cases, weights must be applied after the aggregate command. For example, table 4 contains information on the percent of households that contain at least one child under the age of 15.

This information is not present in the household data file, but it can be created by aggregating the household listing file. The break variables are cluster number and household number. Weights are applied after aggregating because we are interested in the weight percent of households with at least one child under the age of 15, not the weighted number of children under the age of 15 in each household.

You must also check any merge operations if your questionnaire uses case identifiers not present in the standard questionnaire. There are a number of merges in the tabulation programs that will only work if unique identifiers are used.

The Include Command

The SPSS program TABLES.SPS can be used to run all of the tabulation programs at once. It consists of a series of INCLUDE commands that execute the tabulation programs individually. If SPSS encounters an error in a program that is included (i.e., executed by an INCLUDE command), it will immediately stop executing the program and return to the program that included the program (i.e., the program that contained the INCLUDE command).

Because of this, you should only use TABLES.SPS when you have checked and modified all of the individual tabulation programs. Be sure also to remove any INCLUDE command that executes a tabulation program that you are not using.

The INCLUDE command imposes four restrictions on the programs that it executes. The first restriction is that each command must begin in the first column of the program. This restriction appears to limit program indentation, but indented lines can be preceded by the '+' character. The commands below, from the program T16.SPS, illustrate the use of the '+' character to maintain program structure.

```
do if (cage >= 6 and cage <= 9).
+ compute solids = 0.
+ if (BF3G = 1) solids = 100.
end if.
variable labels solids "Solid foods".
```

The second restriction imposed by the INCLUDE command is that if a command continues over multiple lines, column 1 of the continuation lines must be blank. Consider the commands below from the program T1.SPS.

```
add files
  /file=*
  /file='tmp6.sav'.
```

Notice that the subcommands on the second and third lines are indented two columns. (While they need only be indented one column to satisfy the restriction, they have been indented two columns to remain consistent with the MICS programming style.)

The third and fourth restrictions are that command terminators are optional and that an asterisk (*) in the first column of a line indicates a comment line. Neither of these restrictions affects our tabulation programs.

In addition to TABLES.SPS, there is an SPSS program that automates the creation of analysis files for each of the three data entry packages. The programs are named EPIINFO.SPS, IMPS.SPS and ISSA.SPS. These programs should only be used when all of the component programs have been executed and shown to work.

These programs are useful for recreating analysis files when a change is made to one of the file creation programs. They insure that all of the programs will be executed and that they will be executed in the proper order.

Appendix A – Files

File type	Questionnaire	Filename	Subdirectory
Data description			
	Household	MICSHH.DIC	ENTRY
	Women	MICSWM.DIC	ENTRY
	Children	MICSCH.DIC	ENTRY
Data files[1]			
	Household	HH###&.DAT	DATA
	Women	WM###&.DAT	DATA
	Children	CH###&.DAT	DATA
Data entry applications			
	Household	MICSHH.APP	ENTRY
	Women	MICSWM.APP	ENTRY
	Children	MICSCH.APP	ENTRY
Structure check			
	All	CHECK.APP	SUPER
Secondary editing			
	All	EDITING.APP	SUPER
Exporting data			
	Household	MERGEHH.BAT	EXPORT
	Household listing	MERGEHL.BAT	EXPORT
	Women	MERGEWM.BAT	EXPORT
	Children	MERGECH.BAT	EXPORT
Reading data into SPSS			
	Household	READHH.SPS	SPSS
	Household listing	READHL.SPS	SPSS
	Women	READWM.SPS	SPSS
	Children	READCH.SPS	SPSS
Variable and value labels			
	Household	LABELHH.SPS	SPSS
	Household listing	LABELHL.SPS	SPSS
	Women	LABELWM.SPS	SPSS
	Children	LABELCH.SPS	SPSS
Analysis file creation			
	Household listing	MAKEHL.SPS	SPSS
	Women	MAKEWM.SPS	SPSS
	Children	MAKECH.SPS	SPSS
Adding weights			
	All	WEIGHTS.SPS	WEIGHTS
Tabulation[2]			
	All	T??.SPS	SPSS

1.   The # characters in the data file names represent the cluster number; the & character is either an M (main) or a V (verification).
2.   The ? character in the tabulation program names represents the table name.