**Guidelines for Merging Data Files of a MICS Survey**

**MICS data files**

MICS surveys typically collect data using the following standard questionnaires: Household Questionnaire (with Water Quality Testing Questionnaire, when included), Questionnaire for Individual Women age 15-49 years, Questionnaire for Individual Men age 15-49 years, Questionnaire for Children age 5-17 years (MICS6 onwards) and Questionnaire for Children Under Five.

Standard MICS questionnaires allow for different units of analysis. Depending on survey content, up to ten MICS data files are produced and distributed in SPSS format for each survey, corresponding to the units of analysis.

hh.sav - Households

hl.sav - Household members

tn.sav - Mosquito nets in households

wm.sav - Women (15-49 years of age)

bh.sav - Birth history

fg.sav - Female genital mutilation

mm.sav - Maternal mortality

ch.sav - Children under five

fs.sav - Children age 5-17

mn.sav - Men (15-49 years of age)

Each of the data files contains data collected in the relevant questionnaires. Some variables considered to be of critical importance for the analysis are included in all data files, regardless of the questionnaire used to collect the information. For example, some household characteristics that are used as background variables in MICS tabulation plan, such as ethnicity of household head or wealth index, are included in all analysis files.

Table 1 shows the connection between MICS Questionnaires and Modules, and each of the MICS data file.

In cases where all the required information is not present in one single data file, MICS data files need to be combined or merged.

This document describes how MICS data files can be merged and provides examples of SPSS code to do this.

**Table 1. MICS Questionnaires and Modules**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HOUSEHOLD QUESTIONNAIRE | QUESTIONNAIRE FOR INDIVIDUAL WOMENAGE 15-49 YEARS | QUESTIONNAIRE FOR CHILDRENUNDER FIVE | QUESTIONNAIRE FOR INDIVIDUAL MENAGE 15-49 YEARS | QUESTIONNAIRE FOR CHILDRENAGE 5-17 YEARS |
| **hh.sav** | **hl.sav** | **tn.sav** | **wm.sav** | **bh.sav** | **fg.sav** | **mm.sav** | **ch.sav** | **mn.sav** | **fs.sav** |
|
|
| HH | Household Information Panel | HL | List of Household Members | TN | Insecticide Treated Nets | WM | Woman’s Information Panel | BH | Birth History | FG | Female Genital Mutilation | MM | Maternal Mortality | UF | Under-Five Child Information Panel | MWM | Man’s Information Panel | FS | 5-17 Child Information Panel |
| HC | Household Characteristics | ED | Education |  |  | WB | Woman’s Background |  |  |  |  |  |  | UB | Under-Five’s Background | MWB | Man’s Background | CB | Child’s Background |
| ST | Social Transfers |  |  |  |  | MT | Mass Media and ICT |  |  |  |  |  |  | BR | Birth Registration | MMT | Mass Media and ICT | CL | Child Labour |
| EU | Household Energy Use |  |  |  |  | CM | Fertility  |  |  |  |  |  |  | EC | Early Childhood Development | MCM | Fertility | FCD | Child Discipline [5-14] |
| TN | Insecticide Treated Nets |  |  |  |  | DB | Desire for Last Birth |  |  |  |  |  |  | UCD | Child Discipline  | MDV | Attitudes Toward Domestic Violence | FCF | Child Functioning |
| WS | Water and Sanitation |  |  |  |  | MN | Maternal and Newborn Health |  |  |  |  |  |  | UCF | Child Functioning  | MVT | Victimisation | PR | Parental Involvement  |
| HW | Handwashing |  |  |  |  | PN | Post-natal Health Checks |  |  |  |  |  |  | BD | Breastfeeding and Dietary Intake  | MMA | Marriage/Union | FL | Foundational Learning Skills  |
| SA | Salt Iodisation |  |  |  |  | CP | Contraception |  |  |  |  |  |  | IM | Immunisation  | MAF | Adult Functioning  |  |   |
| WQ | Water quality |  |  |  |  | UN | Unmet Need |  |  |  |  |  |  | CA | Care of Illness | MSB | Sexual Behaviour |  |   |
|   |  |  |  |  |  | DV | Attitudes Toward Domestic Violence |  |  |  |  |  |  | AN | Anthropometry | MHA | HIV/AIDS |  |   |
|   |  |  |  |  |  | VT | Victimisation |  |  |  |  |  |  | HF | Vaccination records at health facility  | MMC | Circumcision |  |   |
|   |  |  |  |  |  | MA | Marriage/Union |  |  |  |  |  |  |  |  | MTA | Tobacco and Alcohol Use |  |   |
|   |  |  |  |  |  | AF | Adult Functioning  |  |  |  |  |  |  |  |  | MLS | Life Satisfaction |  |   |
|   |  |  |  |  |  | SB | Sexual Behaviour |  |  |  |  |  |  |  |  |  |  |  |   |
|   |  |  |  |  |  | HA | HIV/AIDS |  |  |  |  |  |  |  |  |  |  |  |   |
|   |  |  |  |  |  | TA | Tobacco and Alcohol Use |  |  |  |  |  |  |  |  |  |  |  |   |
|   |  |  |  |  |  | LS | Life Satisfaction |  |  |  |  |  |  |  |  |  |  |  |   |

**Key variables and relations**

To correctly merge two files, it is necessary to know what variables in the data uniquely identify each case[[1]](#footnote-1). These variables are referred to as key variables. In different MICS files, different combination of multiple variables uniquely identifies each case. For example, each case in the households file (hh.sav) is uniquely identified by cluster number and household number (variables HH1 and HH2 in the hh.sav file). Similarly, each case in the women file (wm.sav) is uniquely identified by cluster number, household number and line number (variables HH1, HH2 and LN in the wm.sav file). The following table shows key variables for MICS data files:

**Table 2. Key Variables for MICS Data Files**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Data file** | **Cluster Number** | **Household Number** | **Household Member Line number** | **Birth History Line Number** | **Daughter Line Number** | **Sibling Line Number** | **Mosquito Net Line Number** |
| hh.sav | HH1 | HH2 |  |  |  |  |  |
| hl.sav | HH1 | HH2 | HL1 |  |  |  |  |
| tn.sav | HH1 | HH2 |  |  |  |  | TNLN |
| wm.sav | HH1 | HH2 | LN |  |  |  |  |
| bh.sav | HH1 | HH2 | LN | BHLN |  |  |  |
| fg.sav | HH1 | HH2 | LN |  | FGLN |  |  |
| mm.sav | HH1 | HH2 | LN |  |  | MMLN |  |
| ch.sav | HH1 | HH2 | LN |  |  |  |  |
| fs.sav | HH1 | HH2 | LN |  |  |  |  |
| mn.sav | HH1 | HH2 | LN |  |  |  |  |

An important step when merging MICS data files is to know the type of relationship between two files, as well as to define desired unit of analysis.

There are two types of relationships: “one to many”, and “one to one”. “One to many” relationship is one where one entity relates to many other. For example, a relationship between households and women, men or children. “One to one” relationship is, for example, a relationship between list of household members and women, men and children.

Desired unit of analysis establishes which one of the data files will be the base (primary) file. That is the data file to which the information from other data files will be appended on.

Merging of data files is based on unique key variables combination that is common to both data files. Key variables, which link the observations of one data file to those of the other, must have the same names in both data files. If names are not the same, renaming of key variables in one of the two data files is required.

This section describes key variables and relations for each separate MICS data file.

1. **hh.sav**

Relations with: hl.sav, tn.sav, wm.sav, bh.sav, fg.sav, mm.sav, ch.sav, fs.sav and mn.sav

Base key variables: HH1 (cluster number) and HH2 (household number)

Instruction to the users:

When merging household members’, women’s, children’s and other data files with their households, you need to use the cluster numbers (variable HH1) and household numbers (variable HH2) as key variables. Since there is a “one-to-many” relationship between households and individuals, you should start with the individual data: household member, women or child, as your “base” (or ‘active data set’) and locate the correct household for each member, meaning that you should be merging the household data sets onto household members’, women's or children’s data, and not the other way around.

1. **hl.sav**

Relations with: wm.sav, bh.sav, fg.sav, mm.sav, ch.sav, fs.sav and mn.sav

Base key variables: HH1 (cluster number), HH2 (household number) and HL1 (member’s line number)

IMPORTANT NOTE: variable HL1 in hl.sav data file is named LN in other files. Renaming of the variable is required prior to merging.

Instruction to the users:

When merging women’s, men’s and children’s data files with household members’ data file you need to use cluster numbers (variable HH1), household numbers (variable HH2) and member’s line number (variable HL1 in household members’ data file and LN in other data files) as key variables. Since many software require key variables to have a unique name, prior to merging data files it is necessary to rename variable HL1 into LN in the hl.sav, or variable LN to HL1 in the wm.sav or ch.sav file or any other file mentioned above. Relationship between household listing and other files listed above is “one-to-one”, meaning that you can merge either hl.sav file to wm.sav or ch.sav file, or the other way around.

1. **wm.sav**

Relations with: bh.sav, fg.sav and mm.sav

Base key variables: HH1 (cluster number), HH2 (household number) and LN (woman line number)

Instruction to the users:

Relationship between wm.sav and other files listed above is “one-to-many”. That means that the “base” (or ‘active data set’) is either bh.sav, fg.sav or mm.sav, and that information from the wm.sav can be merged by using HH1, HH2 and LN as key variables.

As described above, variables from the household data file can be merged to all other data files. Similarly, variables from the household members data file can be merged to women, birth history, female genital mutilation, maternal mortality, children under five, children age 5-17 and men data files. Also, variables from women data file can be appended to birth history, female genital mutilation and maternal mortality data files. In addition to these relationships, few other merging options are possible. For example, variables from women data file can be merged to children under five or children age 5-17 data files. This merge is useful for appending information about the mother or caretaker of a child. Relationship between wm.sav and ch.sav and fs.sav is “one-to-many”, and information from the wm.sav can be merged by using HH1 (cluster number), HH2 (household number) and UF4/FS4 (mothers or caretakers line number in ch.sav and fs.sav, respectively) as key variables. When using software that requires the key variables that are used for merging to have unique names it would be necessary to rename variable LN into UF4 or FS4 in the wm.sav. Important notice is that relationship doesn’t exit for all cases because for some children under five or children age 5-17, women (15-49 years of age) may not be a mother or caretaker. This just one example of the possible scenarios, other similar merging options which may be of interest are not described in detail.

**Step by step example on how to merge hh.sav onto a wm.sav in SPSS**

1. Make sure both files are sorted in ascending order by key variables before trying to merge.



From the menus choose: Data…. Merge Files…. Add Variables...



1. Select the file you wish to merge:

If the file is already open select it from the list of „an open dataset“, and if it is not then browse for the file.



If key variables (HH1 and HH2) have the same name they will show up in the 'Excluded Variables' list. Select the key variables from the 'Excluded Variables' list. Check the 'Match cases on key variables in sorted files' checkbox. Select the 'Both files provide cases' radio button. Now press the right arrow button that is next to the 'Key Variables:' list box.



SPSS will give you a warning regarding sorted key variables. Make sure both files were sorted in ascending order before trying to do a file merge.



And as the final result women’s data file will contain data from household data file.

Alernativly merging the same merging can be done by running the following syntax:

\* open the women file.

get file ="wm.sav".

\* sort cases by ID variables.

sort cases HH1 HH2 LN.

save outfile = "wm.sav".

\* open the household file.

get file ="hh.sav".

\* sort cases by ID variables.

sort cases HH1 HH2.

save outfile = "hh.sav".

\* merge the household data file onto the women file.

match files

 /file = "wm.sav"

 /table = 'hh.sav'

 /by HH1 HH2 .

\*save the women's file.

save outfile = 'wm.sav'.

As mentioned above, when merging hl.sav file to wm.sav or ch.sav files it is important to rename variable HL1 into LN. This action can also be performed by running following syntax:

save outfile = "hl.sav"

/rename = (HL1 = LN).

1. Each household is a case in the household data file, each woman (age 15-49) is a case in the Individual Questionnaire for Women data file, each child age under five is a case in Questionnaire for Children Under Five data file, etc. [↑](#footnote-ref-1)